ON THE RESOLUTION OF MISSPECIFICATION IN STOCHASTIC OPTIMIZATION, VARIATIONAL INEQUALITY, AND GAME-THEORETIC PROBLEMS

BY

HAO JIANG

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy in Industrial Engineering
in the Graduate College of the
University of Illinois at Urbana-Champaign, 2015

Urbana, Illinois

Doctoral Committee:

Associate Professor Angelia Nedich, Chair
Associate Professor Uday V. Shanbhag, Director of Research
Professor Sean P. Meyn
Associate Professor Carolyn L. Beck

Abstract

Traditionally, much of the research in the field of optimization algorithms has assumed that problem parameters are correctly specified. Recent efforts under the robust optimization framework have relaxed this assumption by allowing unknown parameters to vary in a prescribed uncertainty set and by subsequently solving for a worst-case solution. This dissertation considers a rather different approach in which the unknown or misspecified parameter is a solution to a suitably defined (stochastic) learning problem based on having access to a set of samples. Practical approaches in resolving such a set of coupled problems have been either sequential or direct variational approaches. In the case of the former, this entails the following steps: (i) a solution to the learning problem for parameters is first obtained; and (ii) a solution is obtained to the associated parametrized computational problem by using (i). Such avenues prove difficult to adopt particularly since the learning process has to be terminated finitely and consequently, in large-scale or stochastic instances, sequential approaches may often be corrupted by error. On the other hand, a variational approach requires that the problem may be recast as a possibly non-monotone stochastic variational inequality problem; but there are no known first-order (stochastic) schemes currently available for the solution of such problems. Motivated by these challenges, this thesis focuses on studying joint schemes of optimization and learning in three settings: *(i)* misspecified stochastic optimization and variational inequality problems, *(ii)* misspecified stochastic Nash games, *(iii)* misspecified Markov decision processes.

In the first part of this thesis, we present a coupled stochastic approximation scheme which simultaneously solves *both* the optimization and the learning problems. The obtained schemes are shown to be equipped with almost sure convergence properties in regimes when the function f is either strongly convex as well as merely convex. Importantly, the scheme displays the optimal rate for strongly convex problems while in merely convex regimes, through an averaging approach, we quantify the degradation associated with learning by noting that the error in function value after K steps is $\mathcal{O}\left(\sqrt{\ln(K)/K}\right)$, rather than $\mathcal{O}\left(\sqrt{1/K}\right)$ when θ^* is available. Notably, when the averaging window is modified suitably, it can be see that the original rate of $\mathcal{O}\left(\sqrt{1/K}\right)$ is recovered. Additionally, we consider an online counterpart of the misspecified optimization problem and provide a non-asymptotic bound on the average regret with respect to an offline counterpart.

We also extend these statements to a class of stochastic variational inequality problems, an object that unifies stochastic convex optimization problems and a range of stochastic equilibrium problems. Analogous almost-sure convergence statements are provided in strongly monotone and merely monotone regimes, the latter facilitated by using an iterative Tikhonov regularization. In the merely monotone regime, under a weak-sharpness requirement, we quantify the degradation associated with learning and show that expected error associated with $\text{dist}(x_k, X^*)$ is $\mathcal{O}\left(\sqrt{\ln(K)/K}\right)$.

In the second part of this thesis, we present schemes for computing equilibria to two classes of convex stochastic Nash games complicated by a parametric misspecification, a natural concern in the control of large-scale networked engineered system. In both schemes, players *learn the equilibrium strategy* while *resolving the misspecification*: (1) **Stochastic Nash games:** We present a set of coupled stochastic approximation distributed schemes distributed across agents in which the first scheme updates each agent's strategy via a projected (stochastic) gradient step while the second scheme updates every agent's belief regarding its misspecified parameter using an independently specified learning problem. We proceed to show that the produced sequences converge to the true equilibrium strategy and the true parameter in an almost sure sense. Surprisingly, convergence in the equilibrium strategy achieves the *optimal* rate of convergence in a mean-squared sense with a quantifiable degradation in the rate constant; (2) **Stochastic Nash-Cournot games with unobservable aggregate output:** We refine (1) to a Cournot setting where we assume that the tuple of strategies is unobservable while payoff functions and strategy sets are public knowledge through a common knowledge assumption. By utilizing observations of noise-corrupted prices, iterative fixed-point schemes are developed, allowing for *simultaneously* learning the equilibrium strategies and the misspecified parameter in an almost-sure sense.

In the third part of this thesis, we consider the solution of a finite-state infinite horizon Markov Decision Process (MDP) in which both the transition matrix and the cost function are misspecified, the latter in a parametric sense. We consider a data-driven regime in which the learning problem is a stochastic convex optimization problem that resolves misspecification. Via such a framework, we make the following contributions: (1) We first show that a *misspecified* value iteration scheme converges almost surely to its true counterpart and the mean-squared error after K iterations is $\mathcal{O}(1/\sqrt{K})$; (2) An analogous asymptotic almost-sure convergence statement is provided for *misspecified* policy iteration; and (3) Finally, we present a constant steplength *misspecified* Q-learning scheme and show that a suitable error metric is $\mathcal{O}(1/\sqrt{K}) + \mathcal{O}(\sqrt{\delta})$ after K iterations where δ is a bound on the steplength.

To my father and mother

Acknowledgments

This thesis would not have been possible without the support of many people. First of all, I would like to express my sincerest gratitude to my advisor, Prof. Uday Shanbhag. He has always provided me with guidance, support and encouragement as best as he can when I needed not only in the research but also in my life. During my doctoral studies, he taught me the correct way of thinking and how to motivate in the research. Prof. Shanbhag was always very patient to answer my questions, and our detailed and creative discussions often inspired me and let me experienced a memorable time. As a friend, he supported and encouraged me a great deal in my life, especially during the period when I experienced my hardest time of my family. Without his backup, I could not finish the research I have done so far. Also, many thanks to his patience, help and valuable advice during my job search.

I would like to express my deep gratitude to Prof. Sean Meyn. He pointed out the direction for my first research topic in the area of Nash-Cournot game under the setting of the power market. This not only introduced me the world of power markets, but also paved the road for my following research. His intelligent insights and way of thinking inspired me a lot. His passion, rigorous attitude, and commitment to research also made a deep impression in my mind.

I would like to specially thank Prof. Angelia Nedich for being my committee chair. I have learnt a lot from the game theory and variational inequality course offered by her, which provided me a very solid foundation for my research. Her suggestions and comments on my preliminary exam also helped me improve my research. I also thank Carolyn Beck for being in my committee. Her course of data-based systems modeling and her work in the area of Q-learning also inspired me.

And finally, my deep thanks go to my father and mother. They always provided me with unconditional support and encouraged me to overcome difficulties during my study. During my father's last period in his life, he still inquired about my study and provided me with lots of encouragement and hope. Also, my mother, although suffering a lot of pressure, endured the long process of my study and always offered me support and love. Without these, I would not have made the research so far.

Table of Contents

List of Tables . . . . . **viii**

List of Figures . . . . . **ix**

Chapter 1 Introduction **1**
1.1 Misspecified stochastic optimization and variational inequality problems 2
1.2 Misspecified stochastic Nash games 3
1.3 Misspecified Markov decision processes 4
1.4 Notation 4

Chapter 2 Misspecified Stochastic Optimization and Variational Inequality Problems . **6**
2.1 Introduction 6
2.1.1 Related decision-making models 9
2.1.2 Outline and contributions 10
2.2 Stochastic optimization problems with imperfect information 10
2.2.1 Algorithm statement and assumptions 10
2.2.2 Almost-sure convergence 12
2.2.3 Diminishing and constant steplength rate analysis 19
2.2.4 Regret analysis 28
2.3 Stochastic variational inequality problems with imperfect information 33
2.3.1 Almost-sure convergence 33
2.3.2 Diminishing and constant steplength error analysis 40
2.4 Numerical results 44
2.4.1 Problem description 44
2.4.2 Results 45
2.5 Concluding remarks 47

Chapter 3 Misspecified Stochastic Nash Games **49**
3.1 Introduction 49
3.2 Gradient-based schemes for convex Nash games 51
3.2.1 Problem description, assumptions and background 51
3.2.2 Analysis 54
3.3 Iterative fixed-point schemes for misspecified Nash-Cournot games 60
3.3.1 Problem description, assumptions and background 61
3.3.2 Description and definition of algorithm 62
3.3.3 Analysis of noise-corrupted iterative fixed-point schemes 64
3.3.4 Extension to nonlinear price functions 71
3.4 Numerical results 75
3.4.1 Problem description 76
3.4.2 Learning with observation of the aggregate output 77
3.5 Concluding remarks 79

Chapter 4 Misspecified Markov Decision Processes 80
4.1 Introduction 80
4.2 Misspecified value iteration 83
4.3 Misspecified policy iteration 91
4.4 Misspecified Q-learning 94
4.5 Numerical results 99
4.5.1 Problem setting 99
4.5.2 Results 100
4.6 Concluding remarks 101

Chapter 5 Conclusions 102

References 104

List of Tables

2.1 Learning x^* and θ^* in a strongly convex (L) and convex (R) regime: $\xi \sim U[-\theta^*/2, \theta^*/2]$ . . . 46
2.2 Investigation of regret when learning x^* and θ^* in a stochastic convex regime: $\xi \sim U[-\theta^*/2, \theta^*/2]$, $N = 5$, $W = 5$ . . . 47

3.1 Distributed gradient scheme . . . 77
3.2 Iterative fixed-point scheme . . . 77
3.3 Learning x^* and b^* in a stochastic regime when $N = 5$ and $W = 1$, stopping at $k = 10000$ . . . 78

4.1 Misspecified value iteration . . . 100
4.2 Misspecified policy iteration . . . 100
4.3 Misspecified Q-learning . . . 100

List of Figures

2.1 Computing x^* and learning θ^* ($\xi \sim U[-\theta^*/2, \theta^*/2]$, $N = 5$, $W = 5$) 46

3.1 Computing x^* and learning a^* ($\xi \sim U[-\theta^*/2, \theta^*/2]$, $N = 10$) 78
3.2 Computing x^* and learning b^* ($\xi \sim U[-\theta^*/2, \theta^*/2]$, $N = 10$) 78

Chapter 1

Introduction

Increasingly, optimization and game-theoretic problems need to be solved in uncertain and networked regimes complicated by parametric misspecification. One approach relies on estimation of these parameters through a separate learning process that necessitates aggregating data in an offline fashion. Historically, this offline avenue can be formalized by a two-step, and in effect, a serial approach: (i) The first step requires the learning of such parameters by possibly fitting a model to a set of samples, a problem that falls within the purview of statistical learning [1]; (ii) Given an estimate of such parameters, optimization algorithms can be subsequently applied. Unfortunately, in many dynamic settings complicated by streaming data and the need for online decision-making, one cannot impose such a separation in these processes and both optimization and learning need to be carried out simultaneously, particularly when exact solutions to the statistical learning problem can only be obtained in the limit. An alternate approach can be constructed in settings where an offline aggregation of data cannot be managed. Instead, in this setting, the observations are a function of the computational decisions. In this context, we consider an *online* avenue that is customized to the problem of interest (for instance stochastic Nash-Cournot games). Accordingly, in this dissertation, we consider three problem settings corrupted by misspecification in Chapters 2–4:

(i) Static stochastic convex optimization and monotone variational inequality problems;

(ii) Static stochastic Nash games;

(iii) Markov decision processes.

Before proceeding, we provide a short motivation and discussion of the contributions in each of these chapters.

1.1 Misspecified stochastic optimization and variational inequality problems

Convex optimization has proven to be a useful model for resolving a broad class of problems (cf. [2]). In settings where equilibria and competition assume relevance, variational inequality problems have gained immensely in relevance. Yet, in both contexts, it is assumed that the functions (in the context of optimization) and the maps (in the variational inequality setting) are prescribed precisely. However, as problems grow in intricacy and complexity, this assumption cannot be expected to hold. For instance, convex optimization models have found utility in portfolio optimization; however, covariance matrices in such setting rely estimation. Similarly, variational inequality formulations have allowed for capturing imperfectly competitive equilibrium problems; again, the parametrization of the utility functions may not always be available. In short, there is an increasing need to develop algorithms that can resolve misspecification while solving the correctly misspecified problem.

When one considers the joint problem of learning the misspecified parameter and optimizing the system, two approaches may be utilized: (i) The first of these is a sequential approach, i.e. specifying the model and/or parameters based on statistical learning and then solving the resulting optimization problems of interest. Any practically implemented sequential scheme has to terminate the learning problem after finite time. This results in an estimator of the learning problem corrupted by error and this error propagates into the solution of the optimization problem; (ii) A second approach uses the variational avenue and relies on converting the joint learning and optimization problem into a higher dimensional variational inequality problem. However, unless rather strong assumptions are imposed, the mapping associated with the variational inequality problem is not necessarily monotone, which prevents us to use recently developed stochastic approximation schemes for solving monotone stochastic variational inequality problems.

Motivated by the lack of available simultaneous approaches, we propose coupled stochastic approximation schemes in Chapter 2 that allows for solving misspecified stochastic optimization and variational inequality problems. For the misspecified optimization problem, we consider the cases when the function is either strongly convex or merely convex. Almost sure convergence properties can be shown in both cases. When the function is strongly convex, the scheme displays the same optimal rate as the true parameter is available, i.e. $\mathcal{O}\left(\sqrt{1/K}\right)$ after K steps. While in merely convex regimes, we can quantify the degradation associated with learning by using an averaging method, and the error in function value after K steps is $\mathcal{O}\left(\sqrt{\ln(K)/K}\right)$, rather than $\mathcal{O}\left(\sqrt{1/K}\right)$ when parameter information is available. To recover the original rate of $\mathcal{O}\left(\sqrt{1/K}\right)$, we modify the averaging window and get the desired result. In addition, we consider an online counterpart of the misspecified optimization problem and provide a non-asymptotic bound on the average regret. All of

these results can be extended to a class of misspecified stochastic variational inequality problems, which are general cases for stochastic convex optimization and a range of stochastic equilibrium problems. A major difference lies in the merely monotone regimes. We need to use an iterative Tikhonov regularization to get almost-sure convergence results in that case. Also, under merely monotone assumptions, we can quantify the degradation associated with learning and show that the expected distance between the iterate and optimal set is $\mathcal{O}\left(\sqrt{\ln(K)/K}\right)$.

1.2 Misspecified stochastic Nash games

While convex Nash games can be compactly captured by a variational inequality problem, the contributions of the prior section cannot adequately address the intricacies that are presented by Nash games. For instance, a key concern in the computation of equilibria is the need for developing *distributed* protocols that abide by privacy concerns. This motivates the next chapter of this dissertation. In particular, when designing protocols for Nash games, particularly in the absence of a centralized controller, the goal lies computing Nash equilibria when the utility functions are misspecified and rely on agent-specific information that can only be learnt through a set of offline observations. In many regimes, this set of observations may not be available. Consider, for instance, a Nash-Cournot game in which each player decides its own production level of a common commodity while the price of the commodity is based on the aggregate sales. In this regime, players may have a correct model for the price function but an incorrect estimate of its parameters. In this setting, our intent lies in developing an online scheme which relies on observing true prices that allows for learning the misspecified price function parameter. This avenue does not necessitate accumulating observations.

Motivated by these challenges, in Chapter 3, we propose schemes for computing equilibria to misspecified stochastic Nash games. In the proposed schemes, players learn the equilibrium strategy while resolving the misspecification. We consider two settings: (1) general stochastic Nash games with observable aggregate output; (2) stochastic Nash-Cournot games with unobservable aggregate output. In the first case, we propose coupled stochastic approximation distributed schemes across agents. Each agent updates its strategy through a gradient step while updating its belief regarding misspecified parameters through a learning step. Both the true equilibrium strategy and the true parameter can be shown to be achieved in an almost sure sense. The scheme displays the same optimal rate of convergence in the equilibrium strategy in a mean-squared sense as the true parameter is available. In the second case, we consider a special type of Nash games, i.e. Nash-Cournot game, and assume that the aggregate output is unavailable. In addition, we impose a common knowledge assumption: payoff functions and strategy sets are public knowledge. This is a common

assumption for analyzing Nash-Cournot games without information of aggregate output. By using the difference between the observed true price and estimated price, we propose iterative fixed-point schemes which can learn the equilibrium strategies and the misspecified parameter simultaneously in an almost-sure sense. Furthermore, we can extend the result to nonlinear price functions.

1.3 Misspecified Markov decision processes

While the previous two sections have considered static problems, a natural extension lies in sequential decision-making problems. In particular, we consider the Markov decision-making problems (MDPs). Such problems assume relevance in a range of settings (cf. [3, 4]). Yet, in such sense, the transition matrices and the cost functions may be misspecified. Several avenues have been adopted when transition matrices are not known precisely including robust optimization and Q-learning. Yet, there is little available when cost functions are misspecified and in the presence of streaming data, traditional schemes cannot be directly employed. In fact, there is little by way of asymptotics and error analysis for resolving such MDPs with streaming data. Similarly as in misspecified stochastic optimization problems, sequential approaches can, at best, provide approximate solutions.

Motivated by these challenges, we propose a simultaneous scheme for learning and computation in Chapter 4 to solve a finite-state infinite horizon MDP in which the transition matrix and the parametrization of the cost function are unavailable. We consider a data-driven regime in which the learning problem is a stochastic convex optimization problem that resolves misspecification. Three types of schemes are considered: (1) misspecified value iteration scheme; (2) misspecified policy iteration scheme; (3) misspecified Q-learning scheme. The misspecified value iteration scheme can be shown to converge almost surely to its true counterpart and the associated mean-squared error of convergence is provided based on the presence of learning. When the steplength is constant, we can also get an optimized error bound for the value funtion in terms of the number of iteration steps. In the context of misspecified policy iteration scheme, we can provide an analogous asymptotic almost-sure convergence statement and error analysis as in the case with information of the transition matrix and cost function. Finally, we present a constant steplength misspecified Q-learning scheme and provide a suitable error bound based on iteration steps and steplength.

1.4 Notation

Throughout the paper, we use $\|x\|$ to denote the Euclidean norm of a vector x, i.e., $\|x\| = \sqrt{x^T x}$. We use Π_K to denote the Euclidean projection operator onto a set K, i.e., $\Pi_K(x) \triangleq \operatorname{argmin}_{y \in K} \|x - y\|$. A square

matrix H is said to be a $\mathbf{P}$-matrix if every principal minor of H is positive. Similarly, H is a $\mathbf{P_0}$-matrix if every principal minor of H is nonnegative.

Chapter 2

Misspecified Stochastic Optimization and Variational Inequality Problems

2.1 Introduction

In the last two decades, robust optimization [5, 6] approaches have grown in relevance when decision-makers are faced with optimization problems with uncertain parameters. Succinctly, in such an approach, given an uncertainty set that captures the realizations assumed by such a parameter, the *robust* solution represents the *worst-case* over this set of realizations. Naturally, an appropriate choice of such an uncertainty set is crucial and as the availability of data reaches levels hitherto unseen, there is growing interest in data-driven approaches [7] for constructing such sets. Our interest is in closely related yet distinct settings driven by data in which the point estimate of a parameter may be obtained through a learning problem, suitably defined through the aggregation of data. We provide two instances of such problems:

***(i)* Portfolio optimization** Portfolio optimization problems prescribe the optimal constructions of portfolios over a set of assets, for which the mean and covariance of returns are not necessarily known. Traditional approaches have assumed that such returns are available while more recent robust optimization models have utilized factor-based models in constructing uncertainty sets [8, 9, 10]. An alternate, and possibly less conservative, data-driven model of such a problem that employs a point estimate of the mean and covariance matrix requires the solution of two coupled problems: (1) A portfolio optimization problem parametrized by (θ^*, Σ^*) representing the mean and covariance matrix of returns; and (2) A learning problem that utilizes data to obtain the best (θ^*, Σ^*).

***(ii)* Power systems operation** The operation of power grids relies on the solution of hourly (or more frequent) commitment and dispatch problems, each of which is reliant on a range of parameters that are often uncertain. These parameters include supply-side information regarding capacity of wind-power as well as load forecasts. Recently robust optimization approaches have proved to be exceedingly popular [11, 12, 13]. An alternate formulation is given by the following two coupled problems: (1) An economic dispatch problem parametrized by θ^*, a vector that captures the unknown supply and demand side parameters; and (2) A

learning problem that computes θ^* through the accumulation of data.

We believe that such coupled formulations have broad applicability beyond merely the settings mentioned above in (i) and (ii). They may also find application in inventory control problems with stochastic demand [14, 15, 16, 17], robust network design [18], robust routing in communication networks [19], amongst others. To recap the difference between the two problem frameworks, it can be seen that (R-Opt), a robust optimization framework, minimizes the worst-case of the optimal value $f(x;\theta)$ over the uncertainty set $\mathcal{U}_\theta$ while (L-Opt) considers the joint solution of an optimization problem in x, parametrized by θ^*, where θ^* is a solution to a learning problem with a metric $g(\theta)$. The following formulations may provide a clearer comparison:

R-Opt	minimize	$\max_{\theta \in \mathcal{U}_\theta} f(x;\theta)$
	subject to	$x \in X.$

L-Opt	minimize	$f(x;\theta^*)$
	$\underset{x \in X}{\text{minimize}}$	$g(\theta)$
	$\underset{\theta \in \Theta}{\text{minimize}}$	

We consider regimes where the function $f(x;\theta)$ is a convex expected-value function and the resulting problem is given by the following:

$$\min_{x \in X} \ \mathbb{E}[f(x;\theta^*,\xi(\omega))], \tag{$\mathcal{P}_x^o(\theta^*)$}$$

where $X \subseteq \mathbb{R}^n$ is a closed and convex set, $\xi : \Omega \to \mathbb{R}^d$ is a $d-$dimensional random variable defined on a probability space $(\Omega, \mathcal{F}_x, \mathbb{P}_x)$, $f : X \times \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}$ is a real-valued function, and θ^* denotes an $m-$dimensional vector of parameters. Estimating such parameters often requires the resolution of a suitably defined learning problem, given by a stochastic optimization problem ($\mathcal{L}_\theta$), and defined next:

$$\min_{\theta \in \Theta} \quad g(\theta) \triangleq \mathbb{E}[g(\theta;\eta)], \tag{$\mathcal{L}_\theta$}$$

where $\Theta \subseteq \mathbb{R}^m$ is a closed and convex set, $\eta : \Lambda \to \mathbb{R}^p$ is a random variable defined on a probability space $(\Lambda, \mathcal{F}_\theta, \mathbb{P}_\theta)$, and $g : \Theta \times \Lambda \to \mathbb{R}$ is a real-valued function. When one considers the joint problem of learning and optimization, then there are at least two obvious approaches that immediately emerge as possibilities:

(a) Sequential approach: Consider an inherently serial process wherein the first stage incorporates a model/parameter specification phase based on statistical learning while the second stage leverages these findings in developing and solving the actual optimization problem of interest. Such an ordering relies on the learning problems being relatively small and tractable compared to the optimization problems, ensuring that accurate solutions are available within a reasonable time period. Strictly speaking, if one terminates the learning process prematurely with an estimator $\hat{\theta}$, the resulting estimator is essentially corrupted by

error in that $\hat{\theta} \neq \theta^*$. This error propagates into the solution $\hat{x}$ of the computational problem, denoted by $\mathcal{P}_x^o(\hat{\theta})$ and the associated gap might be quite significant. Note that unless the learning problem is solvable via a finite termination algorithm, such a approach cannot provide asymptotic statements but can, at best, provide approximate solutions. Consequently, an inherently serial process reliant on a prematurely truncated learning scheme often fails to provide accurate solutions to the computational problem.

(b) Variational approach: Under suitable convexity and differentiability requirements, the following holds:

$$x^* \text{ solves } (\mathcal{P}_x^o(\theta^*)) \text{ and } \theta^* \text{ solves } (\mathcal{L}_\theta),$$

if and only if (x^*, θ^*) is a solution to the (stochastic) variational inequality problem VI(Z, F) [20] where

$$Z \triangleq X \times \Theta \text{ and } H(z) \triangleq \begin{pmatrix} \mathbb{E}[\nabla_x f(x; \theta, \xi)] \\ \mathbb{E}[\nabla_\theta g(\theta; \eta)] \end{pmatrix}.$$

Recall that z^* is a solution to VI(Z, F) if $(z - z^*)^T F(z) \geq 0$ for all $z \in Z$. Furthermore, if x^* and θ^* denote solutions to $(\mathcal{P}_x^o(\theta^*))$ and $(\mathcal{L}_\theta)$, respectively, then an oft-used avenue in obtaining a solution (x^*, θ^*) entails obtaining a solution to VI(Z, F). However, unless rather strong assumptions are imposed, the map H is not necessarily monotone, precluding the use of recently developed stochastic approximation schemes for solving monotone stochastic variational inequality problems [21, 22, 23], extragradient-based variants [24, 25], and accelerated approaches [26].

Simultaneous approach: This chapter is motivated by the inadequacy of available approaches and, more generally, the absence of *asymptotically convergent schemes with provable non-asymptotic rates*. We present a framework where the learning and the computational problems are solved **simultaneously** via a joint set of stochastic approximation schemes. Such an avenue has several advantages. First, under such an approach, one can provide rigorous statements of asymptotic convergence of the obtained estimators for both, the solution to the computational problem and the associated learning problem. Second, error bounds on the expected error can be provided for a fixed number of steps under a regime with constant and diminishing steplengths. Third, the statements may be extended to the variational regime in which the computational problem is given by the variational counterpart of $(\mathcal{P}_x^o(\theta^*))$, given by $(\mathcal{P}_x^v(\theta^*))$; such a problem requires an $x^* \in X$ such that

$$\mathbb{E}[F(x^*; \theta^*, \xi(\omega))]^T (\, x - x^* \,) \, \geq \, 0, \qquad \forall \, x \, \in \, X, \tag{$\mathcal{P}_x^v(\theta^*)$}$$

where $X \subseteq \mathbb{R}^n$ is a closed and convex set, $\xi : \Omega \to \mathbb{R}^d$ is a $d-$dimensional random variable defined on a

probability space $(\Omega, \mathcal{F}_x, \mathbb{P}_x)$, $F : X \times \mathbb{R}^d \times \mathbb{R}^m \to \mathbb{R}^n$ is a real-valued continuous mapping. Note that when $F(x^*; \theta^*, \xi) \triangleq \nabla_x f(x^*; \theta^*, \xi)$, this reduces to a convex optimization problem. Furthermore, the choice of using a variational problem, rather than merely an optimization problem, is founded on the need to model a variety of multiagent settings complicated by a breadth of strategic interactions, ranging from purely cooperative to distinctly noncooperative [27].

2.1.1 Related decision-making models

While unaware of the availability of general purpose tools that can resolve precisely such problems, we describe settings where such questions have assumed relevance:

Adaptive control [28]: In tracking problems in adaptive control [29], the authors consider a perturbation approach for analyzing a adaptive tracking algorithm and consider three estimation schemes, specifically least mean squares (LMS) scheme, its recursive variant (RLMS), and the Kalman filter (which requires some distributional assumptions on the noise). First, much of this treatment is in the unconstrained regime with tractable (often quadratic estimation objectives), allowing for deriving closed-form (and often linear) update rules. Second, when the noise in the estimation process is Gaussian, the Kalman filter provides a minimum variance estimator. If on the other hand, the noise is non-Gaussian, then the Kalman filter provides the optimal linear estimator (in the sense that no linear filter provides smaller variance). In fact, these assumptions often form the basis of most adaptive control algorithms (cf. [30] and [31] for a discussion adaptive control and stochastic approximation.) Our focus is on static stochastic problems with far less assumptions on the nature of the problem and the associated distributions. Specifically, we allow for more general stochastic convex objectives (or monotone maps in the context of VIs) in either the optimization or the learning problem, allow for convex feasibility sets for both the optimization or the learning problems, and impose relatively mild moment assumptions on the noise (unlike the Gaussian assumptions that are necessary in some of the estimation models).

Iterative learning control: A related avenue lies in iterative learning control (ILC) has its roots in the studies by Uchiyama [32] and Arimoto et al. [33]. ILC [34] is a form of tracking control employed for repetitive control problems, instances being chemical batch processes, robot arm manipulators, and reliability testing rigs. Our problem is more restrictive in its focus (static problems) but allow for more general settings in terms of nonlinearity and the underlying distributional requirements.

Multi-armed bandit problems: The multi-armed bandit (MAB) problem considers the question of how to play given a collection of slot machines faced by a gambler. Each machine provides a random reward from a distribution specific to that machine. The gambler aims to maximize the expected sum of rewards

earned through a sequence of lever pulls. The total discounted reward is maximized by the index policy that pulls the bandit having greatest value of the Gittins index [35]. In effect, the reward function needs to be learnt while optimizing the system. There has been significant research on such problems over the last several decades, including on the question of computation [36] and finite-time analysis [37].

Finally, related questions have also been studied in revenue management where [38] examined the devastating effect of learning with an incorrect model while maximizing revenue.

2.1.2 Outline and contributions

Broadly speaking, this chapter focuses on the development of *stochastic approximation schemes* that generate iterates $\{x_k\}$ and $\{\theta_k\}$ and makes the following contributions. (i) In Section 2.2, we prove the a.s. convergence of the produced iterates to the prescribed solutions and derive error bounds in a standard and an averaging regime. In particular, we quantify the degradation in the convergence rate from introducing an additional learning phase; (ii) Section 2.2 concludes with a precise non-asymptotic bound on the average regret associated with employing the proposed scheme instead of an offline algorithm; (iii) In Section 2.3, we extend the a.s. convergence results to accommodate stochastic variational inequality problems, rather than merely convex optimization problems. Error analysis is carried out under a suitably defined growth property; (iv) In Section 2.4, we provide some supporting numerics and conclude in Section 2.5.

2.2 Stochastic optimization problems with imperfect information

In this section, we focus on examining $(\mathcal{P}_x^o(\theta^*))$ under various assumptions. We begin by stating the coupled stochastic approximation scheme and providing the necessary assumptions in Section 2.2.1. Convergence analysis of the presented scheme is provided in Section 2.2.2 while diminishing and constant steplength rate analysis is performed in Section 2.2.3. We conclude with a discussion of an online algorithm with the associated bounds on the decay of average regret in Section 2.2.4.

2.2.1 Algorithm statement and assumptions

As mentioned in the previous section, we propose a set of coupled stochastic approximation schemes for computing x^* and θ^*.

Algorithm 1 (**Coupled SA schemes for stochastic optimization problems**). **Step 0.** *Given* $x_0 \in X, \theta_0 \in \Theta$ *and sequences* $\{\gamma_{k,x}, \gamma_{k,\theta}\}$, $k := 0$

Step 1.

$$x^{k+1} := \Pi_X \left(x^k - \gamma_{k,x}(\nabla_x f(x^k; \theta^k) + w^k) \right), \qquad k \geq 0 \tag{Opt$_k$}$$

$$\theta^{k+1} := \Pi_\Theta \left(\theta^k - \gamma_{k,\theta}(\nabla_\theta g(\theta^k) + v^k) \right), \qquad k \geq 0 \tag{Learn$_k$}$$

where $w^k \triangleq \nabla_x f(x^k; \theta^k, \xi^k) - \nabla_x f(x^k; \theta^k)$ *and* $v^k \triangleq \nabla_\theta g(\theta^k; \eta^k) - \nabla_\theta g(\theta^k)$.

Step 2. *If* $k > K$, *stop; else* $k := k + 1$, *go to Step. 1.*

We begin by stating an assumption on the functions f and g.

Assumption 1 (Problem properties, A1-1)**.** *Suppose the following hold:*

(i) For every $\theta \in \Theta$, $f(x; \theta)$ *is strongly convex and continuously differentiable with Lipschitz continuous gradients in* x *with convexity constant* μ_x *and Lipschitz constant* L_x, *respectively.*

(ii) For every $x \in X$, *the gradient* $\nabla_x f(x; \theta)$ *is Lipschitz continuous in* θ *with constant* L_θ.

(iii) The function $g(\theta)$ *is strongly convex and continuously differentiable with Lipschitz continuous gradients in* θ *with convexity constant* μ_θ *and Lipschitz constant* C_θ, *respectively.*

Under Assumption (A1-1), the coupled problem admits a unique solution, as shown next.

Lemma 1 (Solvability)**.** *Consider the problems* $(\mathcal{P}_x^o(\theta^*))$ *and* $(\mathcal{L}_\theta)$ *and suppose assumption (A1) holds. Then* $(\mathcal{P}_x^o(\theta^*))$ *and* $(\mathcal{L}_\theta)$ *collectively admit a unique solution.*

Proof. This follows from the strong convexity of g over Θ and the strong convexity of $f(\bullet; \theta)$ over X. ∎

Additionally, we make the following assumptions on the steplength sequences employed in the algorithm.

Assumption 2 (Steplength requirements, A2-1)**.** *Let* $\{\gamma_{k,x}\}$ *and* $\{\gamma_{k,\theta}\}$ *be chosen such that:*

(i) $\sum_{k=0}^{\infty} \gamma_{k,x} = \infty$, $\sum_{k=0}^{\infty} \gamma_{k,x}^2 < \infty$

(ii) $\gamma_{k,\theta} = \gamma_{k,x} L_\theta^2 / (\mu_x \mu_\theta)$.

We define a new probability space $(Z, \mathcal{F}, \mathbb{P})$, where $Z \triangleq \Omega \times \Lambda$, $\mathcal{F} \triangleq \mathcal{F}_x \times \mathcal{F}_\theta$ and $\mathbb{P} \triangleq \mathbb{P}_x \times \mathbb{P}_\theta$. We use $\mathcal{F}_k$ to denote the sigma-field generated by the initial points (x^0, θ^0) and errors (w^l, v^l) for $l = 0, 1, \cdots, k-1$, i.e., $\mathcal{F}_0 = \{(x^0, \theta^0)\}$ and $\mathcal{F}_k = \{(x^0, \theta^0), ((w^l, v^l), l = 0, 1, \cdots, k-1)\}$ for $k \geq 1$. We make the following assumptions on the filtration and errors.

Assumption 3 (A3). *Let the following hold:*

(i) $\mathbb{E}[w^k \mid \mathcal{F}_k] = 0$ *and* $\mathbb{E}[v^k \mid \mathcal{F}_k] = 0$ *a.s. for all* k.

(ii) $\mathbb{E}[\|w^k\|^2 \mid \mathcal{F}_k] \leq \nu_x^2$ *and* $\mathbb{E}[\|v^k\|^2 \mid \mathcal{F}_k] \leq \nu_\theta^2$ *a.s. for all* k.

We conclude this subsection by stating three results (without proof) that will be subsequently employed in developing our convergence statements. The first two of these are relatively well-known super-martingale convergence results (cf. [39, Lemma 10, Pg. 49–50])

Lemma 2. *Let* v_k *be a sequence of nonnegative random variables adapted to* σ*-algebra* $\mathcal{F}_k$ *and such that*

$$\mathbb{E}[v_{k+1}|\mathcal{F}_k] \leq (1 - u_k)v_k + \beta_k \quad \text{for all } k \geq 0 \quad \text{almost surely,}$$

where $0 \leq u_k \leq 1$, $\beta_k \geq 0$, *and* $\sum_{k=0}^{\infty} u_k = \infty$, $\sum_{k=0}^{\infty} \beta_k < \infty$ *and* $\lim_{k\to\infty} \frac{\beta_k}{u_k} = 0$. *Then,* $v_k \to 0$ *a.s.*

Lemma 3. *Let* v_k, u_k, β_k *and* γ_k *be non-negative random variables adapted to* σ*-algebra* $\mathcal{F}_k$. *If* $\sum_{k=0}^{\infty} u_k < \infty$, $\sum_{k=0}^{\infty} \beta_k < \infty$ *and*

$$\mathbb{E}[v_{k+1}|\mathcal{F}_k] \leq (1 + u_k)v_k - \gamma_k + \beta_k \quad \text{for all } k \geq 0 \quad \text{almost surely.}$$

Then, $\{v_k\}$ *is convergent and* $\sum_{k=0}^{\infty} \gamma_k < \infty$ *almost surely.*

Finally, we present a contraction result reliant on monotonicity and Lipschitz continuity requirements (cf. [40, Theorem 12.1.2, Pg. 1109]).

Lemma 4. *Let* $H : K \to \mathbb{R}^n$ *be a mapping that is strongly monotone over* K *with constant* μ, *and Lipschitz continuous over* K *with constant* L. *If* $q \triangleq \sqrt{1 - 2\mu\gamma + \gamma^2 L^2}$, *then for any* $\gamma > 0$, *we have that for any* x, y, *we have* $\|\Pi_K(x - \gamma H(x)) - \Pi_K(y - \gamma H(y))\| \leq q\|x - y\|$.

2.2.2 Almost-sure convergence

Our first convergence result shows that under the prescribed assumptions, Algorithm 1 generates a sequence of iterates that converges to the unique solution.

Proposition 1 (**Almost-sure convergence under strong convexity of** f**).** *Suppose (A1-1), (A2-1) and (A3) hold. Let* $\{x^k, \theta^k\}$ *be computed via Algorithm 1. Then,* $x^k \to x^*$ *and* $\theta^k \to \theta^*$ *a.s. as* $k \to \infty$, *where* θ^* *denotes the unique solution of* $(\mathcal{L}_\theta)$ *and* x^* *denotes the unique solution to* $(\mathcal{P}_x^o(\theta^*))$.

Proof. Note that $x^* = \Pi_X(x^* - \gamma_{k,x}\nabla_x f(x^*;\theta^*))$. Then, by the nonexpansivity of the Euclidean projector, $\|x^{k+1} - x^*\|^2$ may be bounded as follows:

$$\begin{aligned}\|x^{k+1} - x^*\|^2 &= \|\Pi_X(x^k - \gamma_{k,x}(\nabla_x f(x^k;\theta^k) + w^k)) - \Pi_X(x^* - \gamma_{k,x}\nabla_x f(x^*;\theta^*))\|^2 \\ &\leq \|(x^k - x^*) - \gamma_{k,x}(\nabla_x f(x^k;\theta^k) - \nabla_x f(x^*;\theta^*)) - \gamma_{k,x}w^k\|^2.\end{aligned}$$

By adding and subtracting $\gamma_{k,x}\nabla_x f(x^*,\theta^k)$, this expression can be further expanded as follows:

$$\begin{aligned}&\|(x^k - x^*) - \gamma_{k,x}(\nabla_x f(x^k;\theta^k) - \nabla_x f(x^*;\theta^k)) - \gamma_{k,x}(\nabla_x f(x^*;\theta^k) - \nabla_x f(x^*;\theta^*)) - \gamma_{k,x}w^k\|^2 \\ &= \|(x^k - x^*) - \gamma_{k,x}(\nabla_x f(x^k;\theta^k) - \nabla_x f(x^*;\theta^k))\|^2 + \gamma_{k,x}^2\|\nabla_x f(x^*;\theta^k) - \nabla_x f(x^*;\theta^*)\|^2 + \gamma_{k,x}^2\|w^k\|^2 \\ &\quad - 2\gamma_{k,x}[(x^k - x^*) - \gamma_{k,x}(\nabla_x f(x^k;\theta^k) - \nabla_x f(x^*;\theta^k))]^T(\nabla_x f(x^*;\theta^k) - \nabla_x f(x^*;\theta^*)) \\ &\quad - 2\gamma_{k,x}[(x^k - x^*) - \gamma_{k,x}(\nabla_x f(x^k;\theta^k) - \nabla_x f(x^*;\theta^k))]^T w^k + 2\gamma_{k,x}^2(\nabla_x f(x^*;\theta^k) - \nabla_x f(x^*;\theta^*))^T w^k.\end{aligned}$$

By leveraging the fact that $\mathbb{E}[w^k \mid \mathcal{F}_k] = 0$, we have

$$\mathbb{E}[\|x^{k+1} - x^*\|^2 \mid \mathcal{F}_k] \leq \textbf{Term 1} + \textbf{Term 2} + \textbf{Term 3} + \gamma_{k,x}^2\mathbb{E}[\|w^k\|^2 \mid \mathcal{F}_k], \tag{2.1}$$

where **Terms 1 − 3** are defined as follows:

$$\begin{aligned}&\textbf{Term 1} \triangleq \|(x^k - x^*) - \gamma_{k,x}(\nabla_x f(x^k;\theta^k) - \nabla_x f(x^*;\theta^k))\|^2, \\ &\textbf{Term 2} \triangleq \gamma_{k,x}^2\|\nabla_x f(x^*;\theta^k) - \nabla_x f(x^*;\theta^*)\|^2, \\ \text{and } &\textbf{Term 3} \triangleq -2\gamma_{k,x}[(x^k - x^*) - \gamma_{k,x}(\nabla_x f(x^k;\theta^k) - \nabla_x f(x^*;\theta^k))]^T(\nabla_x f(x^*;\theta^k) - \nabla_x f(x^*;\theta^*)).\end{aligned}$$

By Lemma 4 and (A1-1), it follows that

$$\textbf{Term 1} \leq (1 - 2\gamma_{k,x}\mu_x + \gamma_{k,x}^2 L_x^2)\|x^k - x^*\|^2. \tag{2.2}$$

Furthermore, the Lipschitz continuity of $\nabla_x f(x^*;\theta)$ in θ (A1-1) allows for deriving the following bound:

$$\textbf{Term 2} \leq \gamma_{k,x}^2 L_\theta^2\|\theta^k - \theta^*\|^2. \tag{2.3}$$

Finally, **Term 3** can be bounded by invoking the Cauchy-Schwarz inequality, Lemma 4, (A1-1) and the

triangle inequality, we obtain

$$\begin{aligned}
&2\gamma_{k,x}\|(x^k - x^*) - \gamma_{k,x}(\nabla_x f(x^k;\theta^k) - \nabla_x f(x^*;\theta^k))\|\|\nabla_x f(x^*;\theta^k) - \nabla_x f(x^*;\theta^*)\| \\
&\leq 2\gamma_{k,x}\sqrt{1 - 2\gamma_{k,x}\mu_x + \gamma_{k,x}^2 L_x^2}\|x^k - x^*\|L_\theta\|\theta^k - \theta^*\| \\
&\leq 2\gamma_{k,x}L_\theta\|x^k - x^*\|\|\theta^k - \theta^*\| \\
&\leq \gamma_{k,x}\mu_x\|x^k - x^*\|^2 + \gamma_{k,x}(L_\theta^2/\mu_x)\|\theta^k - \theta^*\|^2,
\end{aligned} \tag{2.4}$$

where the last inequality follows from $2a^T b \leq \|a\|^2 + \|b\|^2$. Combining (2.1), (2.2), (2.3) and (2.4), we get

$$\begin{aligned}
\mathbb{E}[\|x^{k+1} - x^*\|^2 \mid \mathcal{F}_k] \leq (1 - \gamma_{k,x}\mu_x + \gamma_{k,x}^2 L_x^2)\|x^k - x^*\|^2 \\
+ (\gamma_{k,x}L_\theta^2/\mu_x + \gamma_{k,x}^2 L_\theta^2)\|\theta^k - \theta^*\|^2 + \gamma_{k,x}^2\nu_x^2.
\end{aligned} \tag{2.5}$$

Recall that θ^* satisfies the fixed point relationship $\theta^* = \Pi_\Theta(\theta^* - \gamma_{\theta,k}\nabla_\theta g(\theta^*))$, which, together with non-expansivity of the Euclidean projector, allows for deriving the following bound on $\|\theta^{k+1} - \theta^*\|^2$:

$$\begin{aligned}
&\|\theta^{k+1} - \theta^*\|^2 = \|\Pi_\Theta(\theta^k - \gamma_{\theta,k}(\nabla_\theta g(\theta^k) + v^k)) - \Pi_\Theta(\theta^* - \gamma_{\theta,k}\nabla_\theta g(\theta^*))\|^2 \\
&\leq \|\theta^k - \theta^* - \gamma_{\theta,k}(\nabla_\theta g(\theta^k) - \nabla_\theta g(\theta^*)) - \gamma_{\theta,k}v^k\|^2 \\
&= \|\theta^k - \theta^* - \gamma_{\theta,k}(\nabla_\theta g(\theta^k) - \nabla_\theta g(\theta^*))\|^2 + \gamma_{\theta,k}^2\|v^k\|^2 - 2(\theta^k - \theta^* - \gamma_{\theta,k}(\nabla_\theta g(\theta^k) - \nabla_\theta g(\theta^*)))^T v^k.
\end{aligned}$$

By taking conditional expectations and by recalling that $\mathbb{E}[v^k \mid \mathcal{F}_k] = 0$, we obtain the following bound:

$$\begin{aligned}
\mathbb{E}[\|\theta^{k+1} - \theta^*\|^2 \mid \mathcal{F}_k] &\leq \|\theta^k - \theta^* - \gamma_{k,\theta}(\nabla_\theta g(\theta^k) - \nabla_\theta g(\theta^*))\|^2 + \gamma_{k,\theta}^2\mathbb{E}[\|v^k\|^2 \mid \mathcal{F}_k] \\
&\leq q_{k,\theta}^2\|\theta^k - \theta^*\|^2 + \gamma_{k,\theta}^2\nu_\theta^2,
\end{aligned} \tag{2.6}$$

where $q_{k,\theta} \triangleq \sqrt{1 - 2\gamma_{k,\theta}\mu_\theta + \gamma_{k,\theta}^2 C_\theta^2}$. Next, by adding (2.5) and (2.6) and by invoking (A2-1), we obtain the following bound.

$$\begin{aligned}
&\mathbb{E}[\|x^{k+1} - x^*\|^2 \mid \mathcal{F}_k] + \mathbb{E}[\|\theta^{k+1} - \theta^*\|^2 \mid \mathcal{F}_k] \\
&\leq (1 - \gamma_{k,x}\mu_x + \gamma_{k,x}^2 L_x^2)\|x^k - x^*\|^2 + (q_{k,\theta}^2 + \gamma_{k,x}L_\theta^2/\mu_x + \gamma_{k,x}^2 L_\theta^2)\|\theta^k - \theta^*\|^2 + \gamma_{k,x}^2\nu_x^2 + \gamma_{k,\theta}^2\nu_\theta^2 \\
&= (1 - \gamma_{k,x}\mu_x + \gamma_{k,x}^2 L_x^2)\|x^k - x^*\|^2 + (1 - \gamma_{k,x}L_\theta^2/\mu_x + \gamma_{k,x}^2(L_\theta^2 + L_\theta^4 C_\theta^2/(\mu_x^2\mu_\theta^2)))\|\theta^k - \theta^*\|^2 \\
&\quad + \gamma_{k,x}^2\nu_x^2 + \gamma_{k,x}^2\nu_\theta^2 L_\theta^4/(\mu_x^2\mu_\theta^2) \\
&\leq (1 - \alpha\gamma_{k,x} + \beta\gamma_{k,x}^2)(\|x^k - x^*\|^2 + \|\theta^k - \theta^*\|^2) + \delta\gamma_{k,x}^2,
\end{aligned}$$

where $\alpha = \min\{\mu_x, L_\theta^2/\mu_x\}$, $\beta = \max\{L_x^2, L_\theta^2 + L_\theta^4 C_\theta^2/(\mu_x^2\mu_\theta^2)\}$ and $\delta = \nu_x^2 + \nu_\theta^2 L_\theta^4/(\mu_x^2\mu_\theta^2)$. From (A2-1), we have that $\sum_{k=0}^{\infty}(\alpha\gamma_{k,x} - \beta\gamma_{k,x}^2) = \infty, \quad \sum_{k=0}^{\infty}\delta\gamma_{k,x}^2 < \infty$, and

$$\lim_{k\to\infty} \frac{\delta\gamma_{k,x}^2}{\alpha\gamma_{k,x} - \beta\gamma_{k,x}^2} = 0.$$

Then, by invoking the super-martingale convergence theorem (Lemma 2), we have that $\|x^k - x^*\|^2 + \|\theta^k - \theta^*\|^2 \to 0$ *a.s.* as $k \to \infty$, which implies that $x^k \to x^*$ and $\theta^k \to \theta^*$ *a.s.* as $k \to \infty$. ∎

Next we weaken the strong convexity requirement on the function f through the following assumption.

Assumption 4 (A1-2)**.** *Suppose the following holds in addition to (A1-1 (ii)) and (A1-1 (iii)).*

(i) For every $\theta \in \Theta$, $f(x;\theta)$ is convex and continuously differentiable with Lipschitz continuous gradients in x with Lipschitz constant L_x.

Furthermore, we make the following assumptions on the steplength sequences employed in the algorithm.

Assumption 5 (A2-2)**.** *Let $\{\gamma_{k,x}\}$, $\{\gamma_{k,\theta}\}$ and some constant $\tau \in (0,1)$ be chosen such that:*

(i) $\sum_{k=0}^{\infty}\gamma_{k,x}^{2-\tau} < \infty$ *and* $\sum_{k=0}^{\infty}\gamma_{k,\theta}^2 < \infty$,

(iii) $\sum_{k=0}^{\infty}\gamma_{k,x} = \infty$ *and* $\sum_{k=0}^{\infty}\gamma_{k,\theta} = \infty$,

(iii) $\beta_k = \frac{\gamma_{k,x}^{\tau}}{2\gamma_{k,\theta}\mu_\theta} \downarrow 0$ *as* $k \to \infty$.

Proceeding as in the previous result, we present a convergence result under these weakened conditions.

Theorem 1 (**Almost-sure convergence under convexity of f**)**.** *Suppose (A1-2), (A2-2) and (A3) hold. Suppose X is bounded and the solution set X^* of $(\mathcal{P}_x^o(\theta^*))$ is nonempty. Let $\{x^k, \theta^k\}$ be computed via Algorithm 1. Then, $\theta^k \to \theta^*$ a.s. as $k \to \infty$, and x^k converges to a random point in X^* a.s. as $k \to \infty$, where θ^* denotes the unique solution of $(\mathcal{L}_\theta)$ and X^* denotes the solution set of $(\mathcal{P}_x^o(\theta^*))$.*

Proof. By the nonexpansivity of the Euclidean projector, we have for any $x^* \in X^*$ that

$$\begin{aligned}\|x^{k+1} - x^*\|^2 &= \|\Pi_X(x^k - \gamma_{k,x}(\nabla_x f(x^k;\theta^k) + w^k)) - \Pi_X(x^*)\|^2 \\ &\leq \|(x^k - x^*) - \gamma_{k,x}\nabla_x f(x^k;\theta^k) - \gamma_{k,x}w^k\|^2.\end{aligned}$$

By adding and subtracting $\gamma_{k,x}\nabla_x f(x^*,\theta^k)$, this expression can be further expanded as follows:

$$\begin{aligned}
&\|(x^k-x^*)-\gamma_{k,x}\nabla_x f(x^k;\theta^*)-\gamma_{k,x}(\nabla_x f(x^k;\theta^k)-\nabla_x f(x^k;\theta^*))-\gamma_{k,x}w^k\|^2\\
&=\|(x^k-x^*)-\gamma_{k,x}\nabla_x f(x^k;\theta^*)\|^2+\gamma_{k,x}^2\|\nabla_x f(x^k;\theta^k)-\nabla_x f(x^k;\theta^*)\|^2+\gamma_{k,x}^2\|w^k\|^2\\
&\quad-2\gamma_{k,x}[(x^k-x^*)-\gamma_{k,x}\nabla_x f(x^k;\theta^*)]^T(\nabla_x f(x^k;\theta^k)-\nabla_x f(x^k;\theta^*))\\
&\quad-2\gamma_{k,x}[(x^k-x^*)-\gamma_{k,x}\nabla_x f(x^k;\theta^*)]^T w^k+2\gamma_{k,x}^2(\nabla_x f(x^k;\theta^k)-\nabla_x f(x^k;\theta^*))^T w^k.
\end{aligned}$$

Noting that $\mathbb{E}[w^k \mid \mathcal{F}_k]=0$, we have

$$\mathbb{E}[\|x^{k+1}-x^*\|^2 \mid \mathcal{F}_k] \le \textbf{Term 1}+\textbf{Term 2}+\textbf{Term 3}+\gamma_{k,x}^2\mathbb{E}[\|w^k\|^2 \mid \mathcal{F}_k], \tag{2.7}$$

where **Terms 1 − 3** are defined as follows:

$$\begin{aligned}
\textbf{Term 1} &\triangleq \|(x^k-x^*)-\gamma_{k,x}\nabla_x f(x^k;\theta^*)\|^2,\\
\textbf{Term 2} &\triangleq \gamma_{k,x}^2\|\nabla_x f(x^k;\theta^k)-\nabla_x f(x^k;\theta^*)\|^2,\\
\text{and } \textbf{Term 3} &\triangleq -2\gamma_{k,x}[(x^k-x^*)-\gamma_{k,x}\nabla_x f(x^k;\theta^*)]^T(\nabla_x f(x^k;\theta^k)-\nabla_x f(x^k;\theta^*)).
\end{aligned}$$

By invoking the convexity of $f(x;\theta)$ in x and the gradient inequality (see A1-2), we have that

$$\begin{aligned}
\textbf{Term 1} &= \|x^k-x^*\|^2+\gamma_{k,x}^2\|\nabla_x f(x^k;\theta^*)\|^2-2\gamma_{k,x}(x^k-x^*)^T\nabla_x f(x^k;\theta^*)\\
&\le \|x^k-x^*\|^2+\gamma_{k,x}^2\|\nabla_x f(x^k;\theta^*)\|^2-2\gamma_{k,x}(f(x^k;\theta^*)-f(x^*;\theta^*))\\
&\le \|x^k-x^*\|^2+2\gamma_{k,x}^2\|\nabla_x f(x^k;\theta^*)-\nabla_x f(x^*;\theta^*)\|^2+2\gamma_{k,x}^2\|\nabla_x f(x^*;\theta^*)\|^2\\
&\quad-2\gamma_{k,x}(f(x^k;\theta^*)-f(x^*;\theta^*)),
\end{aligned}$$

where the last inequality follows from the identity $\|(a-b)+b\|^2 \le 2\|a-b\|^2+2\|b\|^2$. From the Lipschitz continuity of $\nabla_x f(x;\theta)$ in x, the right hand side can be bounded as follows:

$$\begin{aligned}
&\|x^k-x^*\|^2+2\gamma_{k,x}^2\|\nabla_x f(x^k;\theta^*)-\nabla_x f(x^*;\theta^*)\|^2+2\gamma_{k,x}^2\|\nabla_x f(x^*;\theta^*)\|^2-2\gamma_{k,x}(f(x^k;\theta^*)-f(x^*;\theta^*))\\
&\quad\le (1+2\gamma_{k,x}^2L_x^2)\|x^k-x^*\|^2+2\gamma_{k,x}^2\|\nabla_x f(x^*;\theta^*)\|^2-2\gamma_{k,x}(f(x^k;\theta^*)-f(x^*;\theta^*)).
\end{aligned} \tag{2.8}$$

By the Lipschitz continuity of $\nabla_x f(x;\theta)$ in θ (A1-2),

$$\textbf{Term 2} \le \gamma_{k,x}^2L_\theta^2\|\theta^k-\theta^*\|^2. \tag{2.9}$$

By adding and subtracting $\nabla_x f(x^*;\theta^*)$, and by invoking the Lipschitz continuity of $\nabla_x f(x;\theta)$ in x (A1-2) and the triangle inequality, we may derive a bound for **Term 3** as follows:

$$\begin{aligned}
\textbf{Term 3} &\leq 2\gamma_{k,x}\|(x^k - x^*) - \gamma_{k,x}\nabla_x f(x^k;\theta^*)\|\|\nabla_x f(x^k;\theta^k) - \nabla_x f(x^k;\theta^*)\| \\
&\leq 2\gamma_{k,x}\|(x^k - x^*) - \gamma_{k,x}(\nabla_x f(x^k;\theta^*) - \nabla_x f(x^*;\theta^*)) - \gamma_{k,x}\nabla_x f(x^*;\theta^*)\|L_\theta\|\theta^k - \theta^*\| \\
&\leq 2\gamma_{k,x}\left((1+\gamma_{k,x}L_x)\|x^k - x^*\| + \gamma_{k,x}\|\nabla_x f(x^*;\theta^*)\|\right)L_\theta\|\theta^k - \theta^*\| \\
&= 2\gamma_{k,x}L_\theta\|x^k - x^*\|\|\theta^k - \theta^*\| + 2\gamma_{k,x}^2 L_\theta L_x\|x^k - x^*\|\|\theta^k - \theta^*\| + 2\gamma_{k,x}^2 L_\theta\|\nabla_x f(x^*;\theta^*)\|\|\theta^k - \theta^*\|.
\end{aligned}$$

By using the fact that $2ab \leq a^2 + b^2$, we have further that

$$\begin{aligned}
\textbf{Term 3} \leq{}& \gamma_{k,x}^{2-\tau}L_\theta^2\|x^k - x^*\|^2 + \gamma_{k,x}^{\tau}\|\theta^k - \theta^*\|^2 + \gamma_{k,x}^2 L_\theta L_x\|x^k - x^*\|^2 \\
&+ \gamma_{k,x}^2 L_\theta L_x\|\theta^k - \theta^*\|^2 + \gamma_{k,x}^2 L_\theta^2\|\theta^k - \theta^*\|^2 + \gamma_{k,x}^2\|\nabla_x f(x^*;\theta^*)\|^2,
\end{aligned} \tag{2.10}$$

where $\tau \in (0,1)$ is chosen to satisfy (A2-2). Combining (2.7), (2.8), (2.9) and (2.10), we obtain the following bound on the conditional error.

$$\begin{aligned}
\mathbb{E}[\|x^{k+1} - x^*\|^2 \mid \mathcal{F}_k] \leq{}& (1 + \gamma_{k,x}^{2-\tau}L_\theta^2 + \gamma_{k,x}^2(2L_x^2 + L_\theta L_x))\|x^k - x^*\|^2 + (\gamma_{k,x}^{\tau} + \gamma_{k,x}^2(2L_\theta^2 + L_\theta L_x))\|\theta^k - \theta^*\|^2 \\
&+ 3\gamma_{k,x}^2\|\nabla_x f(x^*;\theta^*)\|^2 - 2\gamma_{k,x}(f(x^k;\theta^*) - f(x^*;\theta^*)).
\end{aligned} \tag{2.11}$$

From (2.6), we have that

$$\mathbb{E}[\|\theta^{k+1} - \theta^*\|^2 \mid \mathcal{F}_k] \leq q_{k,\theta}^2\|\theta^k - \theta^*\|^2 + \gamma_{k,\theta}^2\nu_\theta^2, \tag{2.12}$$

where $q_{k,\theta} \triangleq \sqrt{1 - 2\gamma_{k,\theta}\mu_\theta + \gamma_{k,\theta}^2 C_\theta^2}$. Choose $\beta_k = \frac{\gamma_{k,x}^{\tau}}{2\gamma_{k,\theta}\mu_\theta}$ by (A2-2). Note that by assumption $\beta_{k+1} \leq \beta_k$. By multiplying the left hand side of (2.12) by β_{k+1} and adding to the left hand side of (2.11), we get

$$\begin{aligned}
&\mathbb{E}[\|x^{k+1} - x^*\|^2 \mid \mathcal{F}_k] + \beta_{k+1}\mathbb{E}[\|\theta^{k+1} - \theta^*\|^2 \mid \mathcal{F}_k] \leq \mathbb{E}[\|x^{k+1} - x^*\|^2 \mid \mathcal{F}_k] + \beta_k\mathbb{E}[\|\theta^{k+1} - \theta^*\|^2 \mid \mathcal{F}_k] \quad (2.13) \\
&\leq (1 + \gamma_{k,x}^{2-\tau}L_\theta^2 + \gamma_{k,x}^2(2L_x^2 + L_\theta L_x))\|x^k - x^*\|^2 + (\beta_k q_{k,\theta}^2 + \gamma_{k,x}^{\tau} + \gamma_{k,x}^2(2L_\theta^2 + L_\theta L_x))\|\theta^k - \theta^*\|^2 \\
&+ 3\gamma_{k,x}^2\|\nabla_x f(x^*;\theta^*)\|^2 + \beta_k\gamma_{k,\theta}^2\nu_\theta^2 - 2\gamma_{k,x}(f(x^k;\theta^*) - f(x^*;\theta^*)) \\
&\leq (1 + \gamma_{k,x}^{2-\tau}L_\theta^2 + \gamma_{k,x}^2(2L_x^2 + L_\theta L_x))\|x^k - x^*\|^2 + \underbrace{\frac{\beta_k q_{k,\theta}^2 + \gamma_{k,x}^{\tau} + \gamma_{k,x}^2(2L_\theta^2 + L_\theta L_x)}{\beta_k}}_{\textbf{Term 4}} \cdot\beta_k\|\theta^k - \theta^*\|^2 \\
&+ 3\gamma_{k,x}^2\|\nabla_x f(x^*;\theta^*)\|^2 + \beta_k\gamma_{k,\theta}^2\nu_\theta^2 - 2\gamma_{k,x}(f(x^k;\theta^*) - f(x^*;\theta^*)).
\end{aligned}$$

Term 4 on the right hand side of (2.13) can be further expanded as

$$\begin{aligned}\frac{\beta_k q_{k,\theta}^2+\gamma_{k,x}^{\tau}+\gamma_{k,x}^2(2L_\theta^2+L_\theta L_x)}{\beta_k} &= q_{k,\theta}^2+\frac{\gamma_{k,x}^{\tau}+\gamma_{k,x}^2(2L_\theta^2+L_\theta L_x)}{\beta_k}\\ &= 1-2\gamma_{k,\theta}\mu_\theta+\gamma_{k,\theta}^2C_\theta^2+\frac{\gamma_{k,x}^{\tau}}{\beta_k}+\frac{\gamma_{k,x}^2(2L_\theta^2+L_\theta L_x)}{\beta_k} \qquad (2.14)\\ &= 1+\gamma_{k,\theta}^2C_\theta^2+2\gamma_{k,\theta}\gamma_{k,x}^{2-\tau}\mu_\theta(2L_\theta^2+L_\theta L_x).\end{aligned}$$

Combining (2.13) and (2.14), we get

$$\begin{aligned}&\mathbb{E}[\|x^{k+1}-x^*\|^2\mid\mathcal{F}_k]+\beta_{k+1}\mathbb{E}[\|\theta^{k+1}-\theta^*\|^2\mid\mathcal{F}_k]\\ &\le(1+\gamma_{k,x}^{2-\tau}L_\theta^2+\gamma_{k,x}^2(2L_x^2+L_\theta L_x))\|x^k-x^*\|^2+(1+\gamma_{k,\theta}^2C_\theta^2+2\gamma_{k,\theta}\gamma_{k,x}^{2-\tau}\mu_\theta(2L_\theta^2+L_\theta L_x))\beta_k\|\theta^k-\theta^*\|^2\\ &\quad+3\gamma_{k,x}^2\|\nabla_x f(x^*;\theta^*)\|^2+\beta_k\gamma_{k,\theta}^2\nu_\theta^2-2\gamma_{k,x}(f(x^k;\theta^*)-f(x^*;\theta^*))\\ &\le(1+\gamma_{k,\theta}^2C_\theta^2+2\gamma_{k,\theta}\gamma_{k,x}^{2-\tau}\mu_\theta(2L_\theta^2+L_\theta L_x))(\|x^k-x^*\|^2+\beta_k\|\theta^k-\theta^*\|^2)\\ &\quad+(\gamma_{k,x}^{2-\tau}L_\theta^2+\gamma_{k,x}^2(2L_x^2+L_\theta L_x))\|x^k-x^*\|^2\\ &\quad+3\gamma_{k,x}^2\|\nabla_x f(x^*;\theta^*)\|^2+\beta_k\gamma_{k,\theta}^2\nu_\theta^2-2\gamma_{k,x}(f(x^k;\theta^*)-f(x^*;\theta^*)).\end{aligned}$$

We define the following:

$$\begin{aligned}&u_k\triangleq\gamma_{k,\theta}^2C_\theta^2+2\gamma_{k,\theta}\gamma_{k,x}^{2-\tau}\mu_\theta(2L_\theta^2+L_\theta L_x),\sigma_k\triangleq2\gamma_{k,x}(f(x^k;\theta^*)-f(x^*;\theta^*)),\\ &\text{and }\rho_k\triangleq(\gamma_{k,x}^{2-\tau}L_\theta^2+\gamma_{k,x}^2(2L_x^2+L_\theta L_x))\|x^k-x^*\|^2+3\gamma_{k,x}^2\|\nabla_x f(x^*;\theta^*)\|^2+\beta_k\gamma_{k,\theta}^2\nu_\theta^2.\end{aligned}$$

Then, we have

$$\mathbb{E}[\|x^{k+1}-x^*\|^2\mid\mathcal{F}_k]+\beta_{k+1}\mathbb{E}[\|\theta^{k+1}-\theta^*\|^2\mid\mathcal{F}_k]\le(1+u_k)(\|x^k-x^*\|^2+\beta_k\|\theta^k-\theta^*\|^2)+\rho_k-\sigma_k.$$

By boundedness of X and (A2-2), we have that $\sum_{k=0}^{\infty}u_k<\infty$ and $\sum_{k=0}^{\infty}\rho_k<\infty$. So, by Lemma 3 we get that there exists a random variable V such that $\|x^k-x^*\|^2+\beta_k\|\theta^k-\theta^*\|^2\to V$ in an almost sure sense as $k\to\infty$ and $\sum_{k=0}^{\infty}\sigma_k=\sum_{k=0}^{\infty}2\gamma_{k,x}(f(x^k;\theta^*)-f(x^*;\theta^*))<\infty$.

By (A2-2), Lemma 2 and (2.12), we can get that $\|\theta^k-\theta^*\|\to0$ *a.s.* as $k\to\infty$. Thus, it follows that $\|x^k-x^*\|\to V$ *a.s.* as $k\to\infty$. Since $\sum_{k=0}^{\infty}\gamma_{k,x}=\infty$, we get $\liminf_{k\to\infty}f(x^k;\theta^*)=f(x^*;\theta^*)$ *a.s.* as $k\to\infty$. Since the set X is closed, all accumulation points of $\{x^k\}$ lie in X. Furthermore, since $f(x^k;\theta^*)\to f(x^*;\theta^*)$ along a subsequence *a.s.*, by continuity of f it follows that $\{x^k\}$ has a subsequence converging *a.s.* to some point in X, say $\tilde{x}$, which satisfies $f(\tilde{x};\theta^*)=f(x^*;\theta^*)$. That means $\tilde{x}$ is some random point in X^*. Moreover,

since $\|x^k - x^*\|$ is convergent for any $x^* \in X^*$ *a.s.*, the entire sequence $\{x^k\}$ converges to some random point in X^* *a.s.* ∎

2.2.3 Diminishing and constant steplength rate analysis

While the previous section focused on the almost sure convergence of the prescribed learning and computational schemes, a natural question is whether one can develop rate statements. We begin with an examination of the global rate of convergence and show that $\mathcal{O}(1/K)$ rate estimate is derived for an upper bound on the mean-squared error in the solution x_K when $f(\bullet;\theta^*)$ is strongly convex in $(\bullet)$ and K represents the number of steps, consistent with the result obtained for stochastic approximation (cf. [41, 42]). In addition, it is seen that when the function loses strong convexity, an analogous rate estimate is available by using averaging, akin to an approach first employed in [43], where longer stepsizes were suggested with consequent averaging of the obtained iterates.

Proposition 2 (Rate estimates for strongly convex f). *Suppose (A1-1) and (A3) hold. Suppose $\gamma_{x,k} = \lambda_x/k$ and $\gamma_{\theta,k} = \lambda_\theta/k$ with $\lambda_x > 1/\mu_x$ and $\lambda_\theta > 1/(2\mu_\theta)$. Let $\mathbb{E}[\|\nabla_x f(x^k;\theta^k) + w^k\|^2] \leq M^2$ and $\mathbb{E}[\|\nabla_\theta g(\theta^k) + v^k\|^2] \leq M_\theta^2$ for all $x^k \in X$ and $\theta^k \in \Theta$. Let $\{x^k, \theta^k\}$ be computed via Algorithm 1. Then, the following hold after K iterations:*

$$\mathbb{E}[\|\theta^K - \theta^*\|^2] \leq \frac{Q_\theta(\lambda_\theta)}{K} \text{ and } \mathbb{E}[\|x^K - x^*\|^2] \leq \frac{Q_x(\lambda_x)}{K},$$

where $Q_\theta(\lambda_\theta) \triangleq \max\left\{\lambda_\theta^2 M_\theta^2 (2\mu_\theta\lambda_\theta - 1)^{-1}, \mathbb{E}[\|\theta^1 - \theta^*\|^2]\right\}$,

$$Q_x(\lambda_x) \triangleq \max\left\{\lambda_x^2 \widetilde{M}^2 (\mu_x\lambda_x - 1)^{-1}, \mathbb{E}[\|x^1 - x^*\|^2]\right\}, \text{ and } \widetilde{M} \triangleq \sqrt{M^2 + \frac{L_\theta^2 Q_\theta(\lambda_\theta)}{\mu_x\lambda_x}}.$$

Proof. Suppose $A_k \triangleq \frac{1}{2}\|x^k - x^*\|^2$ and $a_k \triangleq \mathbb{E}[A_k]$. Then, A_{k+1} may be bounded as follows by using the non-expansivity of the Euclidean projector:

$$\begin{aligned} A_{k+1} &= \frac{1}{2}\|x^{k+1} - x^*\|^2 = \frac{1}{2}\left\|\Pi_X\left(x^k - \gamma_{x,k}(\nabla_x f(x^k;\theta^k) + w^k)\right) - \Pi_X(x^*)\right\|^2 \\ &\leq \frac{1}{2}\|x^k - x^* - \gamma_{x,k}(\nabla_x f(x^k;\theta^k) + w^k)\|^2 \\ &= A_k + \frac{1}{2}\gamma_{x,k}^2\|\nabla_x f(x^k;\theta^k) + w^k\|^2 - \gamma_{x,k}(x^k - x^*)^T(\nabla_x f(x^k;\theta^k) + w^k). \end{aligned} \tag{2.15}$$

Note that $\mathbb{E}[(x^k-x^*)^T w^k] = \mathbb{E}[\mathbb{E}[(x^k-x^*)^T w^k|\mathcal{F}_k]] = \mathbb{E}[(x^k-x^*)^T \mathbb{E}[w^k|\mathcal{F}_k]] = 0$. By taking expectations on both sides of (2.15) and by invoking the bounds $\mathbb{E}[\|\nabla_x f(x^k;\theta^k)+w^k\|^2] \leq M^2$ and $\mathbb{E}[\|\nabla_\theta g(\theta^k)+v^k\|^2] \leq M_\theta^2$, it follows that

$$a_{k+1} \leq a_k + \frac{1}{2}\gamma_{x,k}^2 M^2 - \gamma_{x,k}\mathbb{E}[(x^k - x^*)^T \nabla_x f(x^k;\theta^k)]. \tag{2.16}$$

But $f(x;\theta)$ is strongly convex in x with constant μ_x for every $\theta \in \Theta$, leading to the following expression:

$$\begin{aligned}\mathbb{E}[(x^k - x^*)^T \nabla_x f(x^k;\theta^k)] &= \mathbb{E}[(x^k - x^*)^T (\nabla_x f(x^k;\theta^k) - \nabla_x f(x^*;\theta^k))] \\ &\quad + \mathbb{E}[(x^k - x^*)^T (\nabla_x f(x^*;\theta^k) - \nabla_x f(x^*;\theta^*))] + \mathbb{E}[(x^k - x^*)^T \nabla_x f(x^*;\theta^*)] \\ &\geq \mu_x \mathbb{E}[\|x^k - x^*\|^2] + \mathbb{E}[(x^k - x^*)^T (\nabla_x f(x^*;\theta^k) - \nabla_x f(x^*;\theta^*))].\end{aligned} \tag{2.17}$$

Combining (2.16) and (2.17), we get

$$\begin{aligned}a_{k+1} &\leq (1 - 2\gamma_{x,k}\mu_x)a_k + \frac{1}{2}\gamma_{x,k}^2 M^2 - \gamma_{x,k}\mathbb{E}[(x^k - x^*)^T (\nabla_x f(x^*;\theta^k) - \nabla_x f(x^*;\theta^*))] \\ &\leq (1 - 2\gamma_{x,k}\mu_x)a_k + \frac{1}{2}\gamma_{x,k}^2 M^2 + \frac{1}{2}\gamma_{x,k}\mu_x \mathbb{E}[\|x^k - x^*\|^2] + \frac{1}{2}\frac{\gamma_{x,k}}{\mu_x}\mathbb{E}[\|\nabla_x f(x^*;\theta^k) - \nabla_x f(x^*;\theta^*)\|^2] \\ &\leq (1 - \gamma_{x,k}\mu_x)a_k + \frac{1}{2}\gamma_{x,k}^2 M^2 + \frac{1}{2}\frac{\gamma_{x,k}}{\mu_x} L_\theta^2 \mathbb{E}[\|\theta^k - \theta^*\|^2].\end{aligned} \tag{2.18}$$

Suppose $\gamma_{\theta,k} = \lambda_\theta/k$. Since the function $g(\theta)$ is strongly convex, we can use the standard rate estimate (cf. inequality (5.292) in [42]) to get the following

$$\mathbb{E}[\|\theta^k - \theta^*\|^2] \leq \frac{Q_\theta(\lambda_\theta)}{k}, \tag{2.19}$$

where $Q_\theta(\lambda_\theta) \triangleq \max\left\{\lambda_\theta^2 M_\theta^2 (2\mu_\theta\lambda_\theta - 1)^{-1}, \mathbb{E}[\|\theta^1 - \theta^*\|^2]\right\}$ with $\lambda_\theta > 1/(2\mu_\theta)$. Suppose $\gamma_{x,k} = \lambda_x/k$, allowing us to claim the following:

$$a_{k+1} \leq \left(1 - \frac{\mu_x\lambda_x}{k}\right)a_k + \frac{1}{2}\frac{\lambda_x^2 M^2}{k^2} + \frac{1}{2}\frac{\lambda_x L_\theta^2 Q_\theta(\lambda_\theta)}{\mu_x k^2} = \left(1 - \frac{\mu_x\lambda_x}{k}\right)a_k + \frac{1}{2}\frac{\lambda_x^2 \widetilde{M}^2}{k^2},$$

where $\widetilde{M} \triangleq \sqrt{M^2 + \frac{L_\theta^2 Q_\theta(\lambda_\theta)}{\mu_x\lambda_x}}$. By assuming that $\lambda_x > 1/\mu_x$, the result follows by observing that

$$\mathbb{E}[\|x^k - x^*\|^2] \leq \frac{Q_x(\lambda_x)}{k},$$

where $Q_x(\lambda_x) \triangleq \max\left\{\lambda_x^2 \widetilde{M}^2 (\mu_x\lambda_x - 1)^{-1}, \mathbb{E}[\|x^1 - x^*\|^2]\right\}$. ∎

Remark: Notice that here we assume that f and g are both smooth and strongly convex. A more general framework is that of composite objectives where the objective is a sume of nonsmooth and smooth stochastic components. Lan [44] proposed the accelerated stochastic approximation (AC-SA) algorithm for solving stochastic composite optimization (SCO) problems and proved that it achieves the optimal rate. In related work, Ghadimi and Lan [45, 46] propose a multi-stage AC-SA algorithm, which possesses an optimal rate of convergence for solving strongly convex SCO problems in terms of the dependence on different problem parameters. While this problem class is beyond the current scope, this approach may aid in refinement of the constants in the Proposition 2 in some regimes.

A shortcoming of the previous result is the need for strong convexity of $f(x, \theta)$ in x for every $\theta \in \Theta$. In our next result, we weaken this requirement and allow for a merely convex f, extending the optimal constant stepsize result in [42]. Specifically, given a prescribed number of iterations, say K, the optimal "constant stepsize" derives the error minimizing steplength; in other words, $\gamma_k = \gamma$ for $1 \leq k \leq K$. This is in contrast with the constant stepsize result presented in Proposition 3, where $\gamma_k = \gamma$ for all k. steps. The following Lipschitzian assumption is imposed on the function $f(x; \theta)$.

Assumption 6 (A6)**.** *Suppose the following holds in addition to (A1-2).*

(i) For every $x \in X$, $f(x; \theta)$ is Lipschitz continuous in θ with constant D_θ.

Theorem 2 (**Rate estimates under convexity of f**)**.** *Suppose (A3) and (A6) hold. Suppose $\mathbb{E}[\|x^k - x^*\|^2] \leq M_x^2$, $\mathbb{E}[\|\nabla_x f(x^k; \theta^k) + w^k\|^2] \leq M^2$ and $\mathbb{E}[\|\nabla_\theta g(\theta^k) + v^k\|^2] \leq M_\theta^2$ for all $x^k \in X$ and $\theta^k \in \Theta$. Let $\{x^k, \theta^k\}$ be computed via Algorithm 1. For $1 \leq i, t \leq k$, we define $v_t \triangleq \frac{\gamma_{x,t}}{\sum_{s=i}^k \gamma_{x,s}}$, $\tilde{x}_{i,k} \triangleq \sum_{t=i}^k v_t x^t$ and $D_X \triangleq \max_{x \in X} \|x - x^1\|$. Suppose for $1 \leq t \leq K$, γ_x is defined as follows:*

$$\gamma_x = \sqrt{\frac{4D_X^2 + L_\theta^2 Q_\theta(\lambda_\theta)(1 + \ln K)}{(M^2 + M_x^2)K}},$$

where $Q_\theta(\lambda_\theta) \triangleq \max\left\{\lambda_\theta^2 M_\theta^2 (2\mu_\theta \lambda_\theta - 1)^{-1}, \mathbb{E}[\|\theta^1 - \theta^*\|^2]\right\}$, and $\gamma_{\theta,k} = \lambda_\theta / K$ with $\lambda_\theta > 1/(2\mu_\theta)$. Then the following holds for $1 \leq i \leq K$:

$$\left|\mathbb{E}[f(\tilde{x}_{i,K}; \theta^K) - f(x^*; \theta^*)]\right| \leq \frac{\sqrt{Q_\theta(\lambda_\theta)} D_\theta + C_{i,K}\sqrt{B_K}}{\sqrt{K}},$$

where $C_{i,K} = \frac{K}{K-i+1}$ and $B_K = (4D_X^2 + L_\theta^2 Q_\theta(\lambda_\theta)(1 + \ln K))(M^2 + M_x^2)$.

Proof. By using the same notation in Proposition 2, we have from (2.16) that

$$\begin{aligned} a_{k+1} &\leq a_k + \frac{1}{2}\gamma_{x,k}^2 M^2 - \gamma_{x,k}\mathbb{E}[(x^k - x^*)^T \nabla_x f(x^k;\theta^k)] \\ &\leq a_k + \frac{1}{2}\gamma_{x,k}^2 M^2 - \gamma_{x,k}\mathbb{E}[(x^k - x^*)^T \nabla_x f(x^k;\theta^*)] - \gamma_{x,k}\mathbb{E}[(x^k - x^*)^T (\nabla_x f(x^k;\theta^k) - \nabla_x f(x^k;\theta^*))]. \end{aligned} \tag{2.20}$$

Note that $f(x;\theta)$ is convex in x for every $\theta \in \Theta$, allowing us to leverage the gradient inequality.

$$\mathbb{E}[(x^k - x^*)^T \nabla_x f(x^k;\theta^*)] \geq \mathbb{E}[f(x^k;\theta^*) - f(x^*;\theta^*)]. \tag{2.21}$$

Combining (2.20) and (2.21), we obtain the following:

$$a_{k+1} \leq a_k + \frac{1}{2}\gamma_{x,k}^2 M^2 - \gamma_{x,k}\mathbb{E}[f(x^k;\theta^*) - f(x^*;\theta^*)] - \gamma_{x,k}\mathbb{E}[(x^k - x^*)^T (\nabla_x f(x^k;\theta^k) - \nabla_x f(x^k;\theta^*))].$$

This allows for constructing the following bounds:

$$\begin{aligned} &\gamma_{x,k}\mathbb{E}[f(x^k;\theta^*) - f(x^*;\theta^*)] \\ &\leq a_k - a_{k+1} + \frac{1}{2}\gamma_{x,k}^2 M^2 - \gamma_{x,k}\mathbb{E}[(x^k - x^*)^T (\nabla_x f(x^k;\theta^k) - \nabla_x f(x^k;\theta^*))] \\ &\leq a_k - a_{k+1} + \frac{1}{2}\gamma_{x,k}^2 M^2 + \frac{1}{2}\gamma_{x,k}^2 \mathbb{E}[\|x^k - x^*\|^2] + \frac{1}{2}\mathbb{E}[\|\nabla_x f(x^k;\theta^k) - \nabla_x f(x^k;\theta^*)\|^2] \\ &\leq a_k - a_{k+1} + \frac{1}{2}\gamma_{x,k}^2 M^2 + \frac{1}{2}\gamma_{x,k}^2 M_x^2 + \frac{1}{2}L_\theta^2 \mathbb{E}[\|\theta^k - \theta^*\|^2] \\ &\leq a_k - a_{k+1} + \frac{1}{2}\gamma_{x,k}^2 (M^2 + M_x^2) + \frac{1}{2}\frac{L_\theta^2 Q_\theta(\lambda_\theta)}{k}, \end{aligned} \tag{2.22}$$

where the second inequality follows from the fact that $2ab \leq a^2 + b^2$, the third inequality follows from the boundedness of $\mathbb{E}[\|x^k - x^*\|^2]$ and Lipschitz continuity of $\nabla_x f(x;\theta)$ in θ, and the last inequality follows from (2.19). As a result, for $1 \leq i \leq k$, we have the following:

$$\begin{aligned} \sum_{t=i}^{k} \gamma_{x,t}\mathbb{E}[f(x^t;\theta^*) - f(x^*;\theta^*)] &\leq \sum_{t=i}^{k}(a_t - a_{t+1}) + \frac{1}{2}\sum_{t=i}^{k}\gamma_{x,t}^2(M^2 + M_x^2) + \frac{1}{2}\sum_{t=i}^{k}\frac{L_\theta^2 Q_\theta(\lambda_\theta)}{t} \\ &\leq a_i + \frac{1}{2}\sum_{t=i}^{k}\gamma_{x,t}^2(M^2 + M_x^2) + \frac{1}{2}\sum_{t=i}^{k}\frac{L_\theta^2 Q_\theta(\lambda_\theta)}{t} \\ &\leq a_i + \frac{1}{2}\sum_{t=i}^{k}\gamma_{x,t}^2(M^2 + M_x^2) + \frac{1}{2}L_\theta^2 Q_\theta(\lambda_\theta)(1 + \ln k). \end{aligned} \tag{2.23}$$

Next, we define $v_t \triangleq \frac{\gamma_{x,t}}{\sum_{s=i}^{k}\gamma_{x,s}}$ and $D_X \triangleq \max_{x\in X}\|x - x^1\|$. The following holds invoking these definitions:

$$\mathbb{E}\left[\sum_{t=i}^{k} v_t f(x^t;\theta^*) - f(x^*;\theta^*)\right] \leq \frac{a_i + \frac{1}{2}\sum_{t=i}^{k}\gamma_{x,t}^2(M^2+M_x^2) + \frac{1}{2}L_\theta^2 Q_\theta(\lambda_\theta)(1+\ln k)}{\sum_{t=i}^{k}\gamma_{x,t}}. \tag{2.24}$$

Next, we consider points given by $\tilde{x}_{i,k} \triangleq \sum_{t=i}^{k} v_t x^t$. By convexity of X, we have that $\tilde{x}_{i,k} \in X$ and by the convexity of $f(x;\theta^*)$ in x, we have $f(\tilde{x}_{i,k};\theta^*) \leq \sum_{t=i}^{k} v_t f(x^t)$. From (2.24) and by noting that $a_1 \leq \frac{1}{2}D_X^2$ and $a_i \leq 2D_X^2$ for $i > 1$, we obtain the following for $1 \leq i \leq k$

$$\mathbb{E}[f(\tilde{x}_{i,k};\theta^*) - f(x^*;\theta^*)] \leq \frac{4D_X^2 + \sum_{t=i}^{k}\gamma_{x,t}^2(M^2+M_x^2) + L_\theta^2 Q_\theta(\lambda_\theta)(1+\ln k)}{2\sum_{t=i}^{k}\gamma_{x,t}}. \tag{2.25}$$

Suppose $\gamma_{x,t} = \gamma_x$ for $t = 1,\ldots,k$. Then, it follows that

$$\mathbb{E}[f(\tilde{x}_{1,k};\theta^*) - f(x^*;\theta^*)] \leq \frac{4D_X^2 + k\gamma_x^2(M^2+M_x^2) + L_\theta^2 Q_\theta(\lambda_\theta)(1+\ln k)}{2k\gamma_x}. \tag{2.26}$$

By minimizing the right hand side in $\gamma_x > 0$, we obtain that

$$\gamma_x = \sqrt{\frac{4D_X^2 + L_\theta^2 Q_\theta(\lambda_\theta)(1+\ln k)}{(M^2+M_x^2)k}}.$$

This implies the following bound:

$$\mathbb{E}[f(\tilde{x}_{1,k};\theta^*) - f(x^*;\theta^*)] \leq \sqrt{\frac{B_k}{k}},$$

where $B_k \triangleq (4D_X^2 + L_\theta^2 Q_\theta(\lambda_\theta)(1+\ln k))(M^2+M_x^2)$. Next, we can also claim that for $1 \leq i \leq k$,

$$\mathbb{E}[f(\tilde{x}_{i,k};\theta^*) - f(x^*;\theta^*)] \leq C_{i,k}\sqrt{\frac{B_k}{k}}, \tag{2.27}$$

where $C_{i,k} = \frac{k}{k-i+1}$. Thus, by employing (2.19), (2.27) and the Lipschitz continuity of $f(x;\theta)$ in θ, we have the required result:

$$\begin{aligned}
\left|\mathbb{E}[f(\tilde{x}_{i,k};\theta^k) - f(x^*;\theta^*)]\right| &\leq \left|\mathbb{E}[f(\tilde{x}_{i,k};\theta^k) - f(\tilde{x}_{i,k};\theta^*)]\right| + \left|\mathbb{E}[f(\tilde{x}_{i,k};\theta^*) - f(x^*;\theta^*)]\right| \\
&\leq D_\theta \mathbb{E}[\|\theta^k - \theta^*\|] + \mathbb{E}[f(\tilde{x}_{i,k};\theta^*) - f(x^*;\theta^*)] \\
&\leq \frac{\sqrt{Q_\theta(\lambda_\theta)}D_\theta}{\sqrt{k}} + \mathbb{E}[f(\tilde{x}_{i,k};\theta^*) - f(x^*;\theta^*)] \leq \frac{\sqrt{Q_\theta(\lambda_\theta)}D_\theta + C_{i,k}\sqrt{B_k}}{\sqrt{k}}.
\end{aligned}$$

∎

Remark: In effect, in the context of learning and optimization, the averaging approach leads to a complexity bound given loosely by

$$\mathcal{O}\left(\frac{a_\theta}{\sqrt{K}} + \underbrace{\frac{b + c_\theta\sqrt{\ln(K)}}{\sqrt{K}}}_{\text{degradation from learning}}\right),$$

where a_θ, b, c_θ are suitably defined. If θ^* is available, then $a_\theta, c_\theta = 0$, leading to the standard bound of $\mathcal{O}(1/\sqrt{K})$. While it is not surprising that the requirement to learn θ^* imposes a degradation, it appears that this degradation is not severe. However, by changing the averaging window, this degradation disappears from a rate standpoint. Specifically, the next result is a corollary of Theorem 2 and uses a modified averaging window, as seen in [41].

Corollary 1 (Rate estimates under convexity of f). *Suppose (A3) and (A6) hold. Suppose $\mathbb{E}[\|x^k - x^*\|^2] \le M_x^2$, $\mathbb{E}[\|\nabla_x f(x^k;\theta^k) + w^k\|^2] \le M^2$ and $\mathbb{E}[\|\nabla_\theta g(\theta^k) + v^k\|^2] \le M_\theta^2$ for all $x^k \in X$ and $\theta^k \in \Theta$. Let $\{x^k, \theta^k\}$ be computed via Algorithm 1. Let k be a positive even number. For $k/2 \le t \le k$, we define $v_t \triangleq \frac{\gamma_{x,t}}{\sum_{s=k/2}^{k}\gamma_{x,s}}$, $\tilde{x}_{k/2,k} \triangleq \sum_{t=k/2}^{k} v_t x^t$ and $D_X \triangleq \max_{x\in X}\|x - x^1\|$. Suppose for $1 \le t \le K$, γ_x is defined as follows:*

$$\gamma_x = \sqrt{\frac{4D_X^2 + L_\theta^2 Q_\theta(\lambda_\theta)(1+\ln 2)}{(M^2 + M_x^2)k}},$$

where $Q_\theta(\lambda_\theta) \triangleq \max\left\{\lambda_\theta^2 M_\theta^2 (2\mu_\theta\lambda_\theta - 1)^{-1}, \mathbb{E}[\|\theta^1 - \theta^*\|^2]\right\}$, and $\gamma_{\theta,k} = \lambda_\theta/K$ with $\lambda_\theta > 1/(2\mu_\theta)$. Then the following holds:

$$\left|\mathbb{E}[f(\tilde{x}_{K/2,K};\theta^K) - f(x^*;\theta^*)]\right| \le \frac{\sqrt{Q_\theta(\lambda_\theta)}D_\theta + 2\sqrt{B}}{\sqrt{K}},$$

where $B \triangleq (4D_X^2 + L_\theta^2 Q_\theta(\lambda_\theta)(1+\ln 2))(M^2 + M_x^2)$.

Proof. When $i = k/2$ where k is a positive even number, the second inequality of (2.23) becomes

$$\begin{aligned}
\sum_{t=k/2}^{k}\gamma_{x,t}\mathbb{E}[f(x^t;\theta^*) - f(x^*;\theta^*)] &\le a_{k/2} + \frac{1}{2}\sum_{t=k/2}^{k}\gamma_{x,t}^2(M^2 + M_x^2) + \frac{1}{2}\sum_{t=k/2}^{k}\frac{L_\theta^2 Q_\theta(\lambda_\theta)}{t} \\
&\le a_{k/2} + \frac{1}{2}\sum_{t=k/2}^{k}\gamma_{x,t}^2(M^2 + M_x^2) + \frac{1}{2}L_\theta^2 Q_\theta(\lambda_\theta)\left[\sum_{t=1}^{k}\frac{1}{t} - \sum_{t=k/2-1}^{k}\frac{1}{t}\right] \\
&\le a_{k/2} + \frac{1}{2}\sum_{t=k/2}^{k}\gamma_{x,t}^2(M^2 + M_x^2) + \frac{1}{2}L_\theta^2 Q_\theta(\lambda_\theta)\left[1 + \ln(k) - \ln(k/2)\right] \\
&\le a_{k/2} + \frac{1}{2}\sum_{t=k/2}^{k}\gamma_{x,t}^2(M^2 + M_x^2) + \frac{1}{2}L_\theta^2 Q_\theta(\lambda_\theta)(1+\ln 2). \qquad (2.28)
\end{aligned}$$

Then, (2.26) becomes

$$\mathbb{E}[f(\tilde{x}_{1,k};\theta^*) - f(x^*;\theta^*)] \leq \frac{4D_X^2 + k\gamma_x^2(M^2 + M_x^2) + L_\theta^2 Q_\theta(\lambda_\theta)(1+\ln 2)}{2k\gamma_x}.$$

By minimizing the right hand side in $\gamma_x > 0$, we obtain that

$$\gamma_x = \sqrt{\frac{4D_X^2 + L_\theta^2 Q_\theta(\lambda_\theta)(1+\ln 2)}{(M^2 + M_x^2)k}}.$$

This implies the following bound:

$$\mathbb{E}[f(\tilde{x}_{1,k};\theta^*) - f(x^*;\theta^*)] \leq \sqrt{\frac{B}{k}},$$

where $B \triangleq (4D_X^2 + L_\theta^2 Q_\theta(\lambda_\theta)(1+\ln 2))(M^2 + M_x^2)$. Next, we can also claim that,

$$\mathbb{E}[f(\tilde{x}_{k/2,k};\theta^*) - f(x^*;\theta^*)] \leq C_k\sqrt{\frac{B}{k}},$$

where $C_k = \frac{k}{k-k/2+1} \leq 2$. Thus, we have the required result:

$$\left|\mathbb{E}[f(\tilde{x}_{k/2,k};\theta^k) - f(x^*;\theta^*)]\right| \leq \frac{\sqrt{Q_\theta(\lambda_\theta)}D_\theta + 2\sqrt{B}}{\sqrt{k}}.$$

∎

We now present a constant steplength error bound where the steplength is fixed over the entire algorithm. As mentioned before, this differs from Theorem 2 in that the number of iterations is not fixed. Constant steplength statements are particularly relevant in networked regimes where the coordination of changing steplength sequences across a collection of agents may prove complicated.

Proposition 3 (Constant steplength error bound). *Suppose (A3) holds. Suppose* $\gamma_{\theta,k} := \gamma_\theta$ *and* $\gamma_{x,k} := \gamma_x$. *Suppose* $\mathbb{E}[\|x^k - x^*\|^2] \le M_x^2$ *and* $\mathbb{E}[\|\nabla_x f(x^k;\theta^k) + w^k\|^2] \le M^2$ *for all* $x^k \in X$. *Suppose* $A_k \triangleq \frac{1}{2}\|x^k - x^*\|^2$ *and* $a_k \triangleq \mathbb{E}[A_k]$. *Let* $\{x^k, \theta^k\}$ *be computed via Algorithm 1.*

(i) Suppose (A1-1) holds. Then, the following holds:

$$\limsup_{k\to\infty} a_k \le \frac{1}{2\mu_x}\gamma_x M^2 + \frac{1}{2\mu_x^2}\frac{\gamma_\theta \nu_\theta^2 L_\theta^2}{2\mu_\theta - \gamma_\theta C_\theta^2};$$

(ii) Suppose (A1-2) and (A6) hold and $0 < \tau < 1$. *Then, the following holds:*

$$\limsup_{k\to\infty} \left|\mathbb{E}[f(x^k;\theta^k) - f(x^*;\theta^*)]\right| \le \frac{1}{2}\gamma_x M^2 + \frac{1}{2}\gamma_x^{1-\tau} M_x^2 + \frac{1}{2}\gamma_x^{\tau-1} L_\theta^2 \frac{\gamma_\theta \nu_\theta^2}{2\mu_\theta - \gamma_\theta C_\theta^2} + D_\theta \sqrt{\frac{\gamma_\theta \nu_\theta^2}{2\mu_\theta - \gamma_\theta C_\theta^2}}.$$

Proof. By (2.6), we get the following:

$$\mathbb{E}[\|\theta^{k+1} - \theta^*\|^2 \mid \mathcal{F}_k] \le q_{k,\theta}^2 \|\theta^k - \theta^*\|^2 + \gamma_{k,\theta}^2 \nu_\theta^2,$$

where $q_{k,\theta} \triangleq \sqrt{1 - 2\gamma_{k,\theta}\mu_\theta + \gamma_{k,\theta}^2 C_\theta^2}$. Suppose $\gamma_{\theta,k} := \gamma_\theta$ is chosen such that $(1 - q_\theta) < 1$ where $q_{\theta,k} := q_\theta$. By taking the expectation and limit supremum on both sides, we have

$$\limsup_{k\to\infty} \mathbb{E}[\|\theta^{k+1} - \theta^*\|^2] \le q_\theta^2 \limsup_{k\to\infty} \mathbb{E}[\|\theta^k - \theta^*\|^2] + \gamma_\theta^2 \nu_\theta^2,$$

or,

$$\limsup_{k\to\infty} \mathbb{E}[\|\theta^k - \theta^*\|^2] \le \frac{\gamma_\theta \nu_\theta^2}{2\mu_\theta - \gamma_\theta C_\theta^2}. \tag{2.29}$$

(i) f is strongly convex: From (2.18), for $\gamma_{x,k} := \gamma_x$ where γ_x is sufficiently small, we have the following:

$$a_{k+1} \le (1 - \gamma_x \mu_x) a_k + \frac{1}{2}\gamma_x^2 M^2 + \frac{1}{2}\frac{\gamma_x}{\mu_x} L_\theta^2 \mathbb{E}[\|\theta^k - \theta^*\|^2].$$

It follows that

$$\begin{aligned}\limsup_{k\to\infty} a_{k+1} &\le (1 - \gamma_x\mu_x) \limsup_{k\to\infty} a_k + \frac{1}{2}\gamma_x^2 M^2 + \frac{1}{2}\frac{\gamma_x}{\mu_x} L_\theta^2 \limsup_{k\to\infty} \mathbb{E}[\|\theta^k - \theta^*\|^2] \\ &\le (1 - \gamma_x\mu_x) \limsup_{k\to\infty} a_k + \frac{1}{2}\gamma_x^2 M^2 + \frac{1}{2}\frac{\gamma_x}{\mu_x} L_\theta^2 \frac{\gamma_\theta \nu_\theta^2}{2\mu_\theta - \gamma_\theta C_\theta^2}.\end{aligned}$$

It follows that

$$\limsup_{k\to\infty} a_k \le \frac{1}{2\mu_x}\gamma_x M^2 + \frac{1}{2}\frac{1}{\mu_x^2} L_\theta^2 \frac{\gamma_\theta \nu_\theta^2}{2\mu_\theta - \gamma_\theta C_\theta^2}.$$

(ii) f is convex: From (2.22), for $\gamma_{x,k} := \gamma_x$, we have the following:

$$\begin{aligned}
\gamma_x \mathbb{E}[f(x^k;\theta^*) - f(x^*;\theta^*)] &\le a_k - a_{k+1} + \frac{1}{2}\gamma_x^2 M^2 - \gamma_x \mathbb{E}[(x^k - x^*)^T(\nabla_x f(x^k;\theta^k) - \nabla_x f(x^k;\theta^*))] \\
&\le a_k - a_{k+1} + \frac{1}{2}\gamma_x^2 M^2 + \frac{1}{2}\gamma_x^{2-\tau}\mathbb{E}[\|x^k - x^*\|^2] \\
&\quad + \frac{1}{2}\gamma_x^{\tau}\mathbb{E}[\|\nabla_x f(x^k;\theta^k) - \nabla_x f(x^k;\theta^*)\|^2] \\
&\le a_k - a_{k+1} + \frac{1}{2}\gamma_x^2 M^2 + \frac{1}{2}\gamma_x^{2-\tau} M_x^2 + \frac{1}{2}\gamma_x^{\tau} L_\theta^2 \mathbb{E}[\|\theta^k - \theta^*\|^2],
\end{aligned}$$

where $0 < \tau < 1$. It follows that

$$\begin{aligned}
\gamma_x \limsup_{k\to\infty} \mathbb{E}[f(x^k;\theta^*) - f(x^*;\theta^*)] &\le \limsup_{k\to\infty} a_k - \limsup_{k\to\infty} a_{k+1} + \frac{1}{2}\gamma_x^2 M^2 + \frac{1}{2}\gamma_x^{2-\tau} M_x^2 \\
&\quad + \frac{1}{2}\gamma_x^{\tau} L_\theta^2 \limsup_{k\to\infty} \mathbb{E}[\|\theta^k - \theta^*\|^2] \\
&\le \frac{1}{2}\gamma_x^2 M^2 + \frac{1}{2}\gamma_x^{2-\tau} M_x^2 + \frac{1}{2}\gamma_x^{\tau} L_\theta^2 \frac{\gamma_\theta \nu_\theta^2}{2\mu_\theta - \gamma_\theta C_\theta^2}.
\end{aligned}$$

It follows that

$$\limsup_{k\to\infty} \mathbb{E}[f(x^k;\theta^*) - f(x^*;\theta^*)] \le \frac{1}{2}\gamma_x M^2 + \frac{1}{2}\gamma_x^{1-\tau} M_x^2 + \frac{1}{2}\gamma_x^{\tau-1} L_\theta^2 \frac{\gamma_\theta \nu_\theta^2}{2\mu_\theta - \gamma_\theta C_\theta^2}.$$

By the Lipschitz continuity of $f(x;\theta)$ in θ (A6(i)), Hölder's inequality and (2.29), we have

$$\begin{aligned}
\limsup_{k\to\infty} \left|\mathbb{E}[f(x^k;\theta^k) - f(x^k;\theta^*)]\right| &\le D_\theta \limsup_{k\to\infty} \mathbb{E}[\|\theta^k - \theta^*\|] \\
&\le D_\theta \limsup_{k\to\infty} \sqrt{\mathbb{E}[\|\theta^k - \theta^*\|^2]} \\
&= D_\theta \sqrt{\limsup_{k\to\infty} \mathbb{E}[\|\theta^k - \theta^*\|^2]} \\
&\le D_\theta \sqrt{\frac{\gamma_\theta \nu_\theta^2}{2\mu_\theta - \gamma_\theta C_\theta^2}}.
\end{aligned}$$

Therefore,

$$\limsup_{k\to\infty} \left|\mathbb{E}[f(x^k;\theta^k) - f(x^*;\theta^*)]\right| \leq \limsup_{k\to\infty} \left|\mathbb{E}[f(x^k;\theta^k) - f(x^k;\theta^*)]\right| + \limsup_{k\to\infty} \left|\mathbb{E}[f(x^k;\theta^*) - f(x^*;\theta^*)]\right|$$
$$\leq \frac{1}{2}\gamma_x M^2 + \frac{1}{2}\gamma_x^{1-\tau} M_x^2 + \frac{1}{2}\gamma_x^{\tau-1} L_\theta^2 \frac{\gamma_\theta \nu_\theta^2}{2\mu_\theta - \gamma_\theta C_\theta^2} + D_\theta \sqrt{\frac{\gamma_\theta \nu_\theta^2}{2\mu_\theta - \gamma_\theta C_\theta^2}}.$$

∎

2.2.4 Regret analysis

In this subsection, we consider the problem of online convex programming in a misspecified regime. In online convex programming problems, a decision-maker sees an infinite sequence of functions $c_1, c_2, \ldots$ where each function is convex in its argument over a closed and convex set X. An *online* convex programming algorithm [47] generates an iterate x_k at each time epoch k and a metric of performance is the *regret* associated with not using an offline algorithm that considers the following problem: $\min_{x\in X} \quad \sum_{k=1}^{K} c_k(x)$. If an online convex algorithm generates iterates $x_1, x_2, \ldots$, then the regret R_K is defined as

$$R_K \triangleq \left[\sum_{k=1}^{K} c_k(x_k) - \min_{x\in X} \sum_{k=1}^{K} c_k(x)\right].$$

A desirable feature of an online convex programming algorithm is that it is characterized by sublinear regret [47], which is given by the following theorem.

Theorem 3 (Theorem 1 in [47]). *Select an arbitrary $x^1 \in F$ and a sequence of learning rates $\eta_1, \eta_2, \ldots \in \mathbb{R}^+$. In time step t, after receiving a cost function, select the next vector x^{t+1} according to the Greedy Projection algorithm:*

$$x^{t+1} = \Pi_F(x^t - \eta_t \nabla c^t(x^t)).$$

If $\eta_t = t^{-1/2}$, the regret of the Greedy Projection algorithm is:

$$R_G(T) \leq \frac{\|F\|^2 \sqrt{T}}{2} + \left(\sqrt{T} - \frac{1}{2}\right) \|\nabla c\|^2,$$

where $\|F\| \triangleq \max_{x,y\in F} d(x,y)$ and $\|\nabla c\| \triangleq \max_{x\in F, t\in\{1,2,\ldots\}} \|\nabla c^t(x)\|$.

Proof sketch: The regret of the Greedy Projection algorithm can be bounded as follows:

$$R_G(T) \leq \frac{\|F\|^2}{2\eta_T} + \frac{\|\nabla c\|^2}{2} \sum_{t=1}^{T} \eta_t.$$

The result can be immediately obtained when $\eta_t = t^{-1/2}$. ∎

Often the model prescribed in an online optimization regime can be refined to a setting where the functions are related across time rather than being a sequence of unrelated functions. We consider one particular regime in which the decision-maker sees a sequence of functions given by $f(\bullet;\theta_1)$, $f(\bullet;\theta_2), \ldots$. Furthermore, neither the values $\theta_1, \theta_2, \ldots$ are known to the decision-maker nor is the fact that $\theta_k \to \theta^*$ as $k \to \infty$. As earlier, we assume that the decision-maker has to furnish $x_1, x_2, \ldots$ and we define the *misspecified regret* after K steps associated with our generated sequence $\{x^k, \theta^k\}$ as follows:

$$R_K \triangleq \mathbb{E}\left[\sum_{k=1}^{K} f(x^k;\theta^k,\xi) - K f(x^*;\theta^*,\xi)\right].$$

Unlike the traditional definition, we consider the departure from $f(x^*,\theta^*)$ and should be contrasted with the standard regret metric given by $R_K^{\text{std}} \triangleq \mathbb{E}\left[\sum_{k=1}^{K} f(x^k;\theta^*,\xi) - K f(x^*;\theta^*,\xi)\right]$. For purposes of deriving analytical bounds, we define the following variant of regret as follows:

$$\widehat{R}_K \triangleq \mathbb{E}\left[\sum_{k=1}^{K} f(x^k;\theta^k,\xi) - \sum_{k=1}^{K} f(y_K^*;\theta^k,\xi)\right], \text{ where } y_K^* \triangleq \operatorname*{argmin}_{y\in X} \mathbb{E}\left[\sum_{k=1}^{K} f(y;\theta^k,\xi)\right].$$

Next, we provide a rate of decay of the upper bound of average regret.

Theorem 4 (**Regret under convexity of f**). *Suppose (A3) and (A6) hold. Suppose* $\mathbb{E}[\|x - x^*\|^2] \leq M_x^2$, $\mathbb{E}[\|\nabla_x f(x;\theta) + w^k\|^2] \leq M^2$ *and* $\mathbb{E}[\|\nabla_\theta g(\theta) + v^k\|^2] \leq M_\theta^2$ *for all* $x \in X$ *and* $\theta \in \Theta$. *Suppose* $\mathbb{E}[\|\nabla_x f(y_K^*;\theta^k) + u^k\|^2] \leq M^2$, *where* $u^k \triangleq \mathbb{E}[\nabla_x f(y_K^*;\theta^k,\xi)] - \nabla_x f(y_K^*;\theta^k)$. *Let* $\{x^k,\theta^k\}$ *be computed via Algorithm 1. Suppose* $\gamma_{k,x} = k^{-\alpha}$ *with* $0.5 \leq \alpha < 1$, *and* $\gamma_{\theta,k} = \lambda_\theta/k$ *with* $\lambda_\theta > 1/(2\mu_\theta)$. *If* $0 < \beta < 1$, *then the following holds:*

$$\frac{R_K}{K} \leq \frac{M_x^2 K^{\alpha-1}}{2} + \frac{M^2(K^{1-\alpha}-\alpha)}{2(1-\alpha)K} + \frac{D_\theta\sqrt{Q_\theta(\lambda_\theta)}(2\sqrt{K}-1)}{K} + \frac{M_x^2}{2K^\beta} + \frac{L_\theta^2 Q_\theta(\lambda_\theta)(\ln(K)+1)}{2K^{1-\beta}},$$

where $\beta > 0$. *Furthermore,*

$$\limsup_{K\to\infty} \frac{R(K)}{K} \leq 0.$$

Proof. By using the proof in Theorem 1 in [47] (cf. Theorem 3), we obtain that $\widehat{R}_K/K$ is bounded as follows:

$$\widehat{R}_K \leq \frac{M_x^2}{2\gamma_{K,x}} + \frac{M^2}{2}\sum_{k=1}^K \gamma_{k,x}.$$

Next, if $\gamma_{k,x} = k^{-\alpha}$ with $0.5 \leq \alpha < 1$, then we have the following bound on $\sum_{k=1}^K \gamma_{k,x}$:

$$\sum_{k=1}^K \gamma_{k,x} = \sum_{k=1}^K k^{-\alpha} \leq 1 + \int_1^K x^{-\alpha}dx = \frac{1}{1-\alpha}(K^{1-\alpha}-\alpha).$$

Therefore, we obtain the following bound on $\widehat{R}_K$:

$$\widehat{R}_K \leq \frac{M_x^2 K^\alpha}{2} + \frac{M^2(K^{1-\alpha}-\alpha)}{2(1-\alpha)}. \tag{2.30}$$

Recall that the difference between the real regret and misspecified regret is given by the following:

$$\begin{aligned}\left|R_K - \widehat{R}_K\right| &= \left|\mathbb{E}\left[\sum_{k=1}^K f(y_K^*;\theta^k,\xi) - Kf(x^*;\theta^*,\xi)\right]\right| \\ &\leq \left|\mathbb{E}\left[\sum_{k=1}^K f(y_K^*;\theta^k,\xi) - Kf(y_K^*;\theta^*,\xi)\right]\right| + \left|\mathbb{E}\left[K\left(f(y_K^*;\theta^*,\xi) - f(x^*;\theta^*,\xi)\right)\right]\right|,\end{aligned}$$

or

$$\frac{\left|R_K - \widehat{R}_K\right|}{K} \leq \underbrace{\left|\mathbb{E}\left[\frac{1}{K}\sum_{k=1}^{K} f(y_K^*;\theta^k,\xi) - f(y_K^*;\theta^*,\xi)\right]\right|}_{\textbf{Term 1}} + \underbrace{\left|\mathbb{E}\left[f(y_K^*;\theta^*,\xi) - f(x^*;\theta^*,\xi)\right]\right|}_{\textbf{Term 2}}. \tag{2.31}$$

We proceed to derive bounds for **Terms** 1 and 2. **Term 1** in (2.31) may be bounded as follows:

$$\begin{aligned}\left|\mathbb{E}\left[\frac{1}{K}\sum_{k=1}^{K} f(y_K^*;\theta^k,\xi) - f(y_K^*;\theta^*,\xi)\right]\right| &\leq \frac{1}{K}\sum_{k=1}^{K}\mathbb{E}\left[\left|f(y_K^*;\theta^k,\xi) - f(y_K^*;\theta^*,\xi)\right|\right] \\ &\leq \frac{D_\theta}{K}\sum_{k=1}^{K}\mathbb{E}[\|\theta^k - \theta^*\|] \\ &\leq \frac{D_\theta}{K}\sum_{k=1}^{K}\sqrt{\frac{Q_\theta(\lambda_\theta)}{k}}.\end{aligned}$$

where the second and third inequalities follow from the Lipschitz continuity of $\nabla f(y^*;\theta)$ in θ (A6) and (2.19). Through some analysis, the right hand side may be further bounded as follows:

$$\frac{D_\theta}{K}\sum_{k=1}^{K}\sqrt{\frac{Q_\theta(\lambda_\theta)}{k}} \leq \frac{D_\theta\sqrt{Q_\theta(\lambda_\theta)}}{K}\left(1+\int_1^K \frac{1}{\sqrt{x}}dx\right) \leq \frac{D_\theta\sqrt{Q_\theta(\lambda_\theta)}(2\sqrt{K}-1)}{K}. \tag{2.32}$$

This implies that **Term 1** in (2.31) converges to zero as $K \to \infty$. Next, we consider **Term 2** in (2.31). By the optimality condition for y_K^*, we have the following expression:

$$\begin{aligned}0 &\geq \sum_{k=1}^{K}\mathbb{E}[(y_K^* - x^*)^T\nabla_x f(y_K^*;\theta^k,\xi)] \\ &= \sum_{k=1}^{K}\mathbb{E}[(y_K^* - x^*)^T\nabla_x f(y_K^*;\theta^*,\xi)] + \sum_{k=1}^{K}\mathbb{E}[(y_K^* - x^*)^T(\nabla_x f(y_K^*;\theta^k,\xi) - \nabla_x f(y_K^*;\theta^*,\xi))].\end{aligned} \tag{2.33}$$

Since $f(x;\theta)$ is convex in x for every $\theta \in \Theta$, we may leverage the gradient inequality.

$$\begin{aligned}&\mathbb{E}[f(x^*;\theta^*,\xi)] \geq \mathbb{E}[f(y_K^*;\theta^*,\xi)] + \mathbb{E}[\nabla_x f(y_K^*;\theta^*,\xi)^T(x^* - y_K^*)] \\ \implies &\mathbb{E}[(y_K^* - x^*)^T\nabla_x f(y_K^*;\theta^*,\xi)] \geq \mathbb{E}[f(y_K^*;\theta^*,\xi) - f(x^*;\theta^*,\xi)].\end{aligned} \tag{2.34}$$

Combining (2.33) and (2.34), we get the following lower bound:

$$0 \geq \sum_{k=1}^{K}\mathbb{E}[f(y_K^*;\theta^*,\xi) - f(x^*;\theta^*,\xi)] + \sum_{k=1}^{K}\mathbb{E}[(y_K^* - x^*)^T(\nabla_x f(y_K^*;\theta^k,\xi) - \nabla_x f(y^*;\theta^*,\xi))].$$

This allows for constructing the following bound on $\sum_{k=1}^{K} \mathbb{E}[f(y_K^*;\theta^*,\xi) - f(x^*;\theta^*,\xi)]$:

$$\begin{aligned}
\sum_{k=1}^{K} \mathbb{E}[f(y_K^*;\theta^*,\xi) - f(x^*;\theta^*,\xi)] &\leq -\sum_{k=1}^{K} \mathbb{E}[(y_K^* - x^*)^T(\nabla_x f(y_K^*;\theta^k,\xi) - \nabla_x f(y_K^*;\theta^*,\xi))] \\
&\leq \frac{1}{2}\sum_{k=1}^{K} \delta_K \mathbb{E}[\|y_K^* - x^*\|^2] + \frac{1}{2}\sum_{k=1}^{K} \frac{1}{\delta_K}\mathbb{E}[\|\nabla_x f(y_K^*;\theta^k,\xi) - \nabla_x f(y_K^*;\theta^*,\xi)\|^2] \\
&\leq \frac{1}{2}\sum_{k=1}^{K} \delta_K M_x^2 + \frac{1}{2}\sum_{k=1}^{K} \frac{1}{\delta_K} L_\theta^2 \mathbb{E}[\|\theta^k - \theta^*\|^2] \\
&\leq \frac{1}{2}\sum_{k=1}^{K} \delta_K M_x^2 + \frac{1}{2}\sum_{k=1}^{K} \frac{1}{\delta_K} \frac{L_\theta^2 Q_\theta(\lambda_\theta)}{k}, \qquad (2.35)
\end{aligned}$$

where $\delta_K = K^{-\beta}$ with $0 < \beta < 1$ and the last inequality follows from (2.19). Note that $\sum_{k=1}^{K} \frac{1}{k} \leq \ln(K)+1$.

$$\begin{aligned}
\text{Thus, } |\mathbb{E}\left[f(y_K^*;\theta^*,\xi) - f(x^*;\theta^*,\xi)\right]| &= \mathbb{E}[f(y_K^*;\theta^*) - f(x^*;\theta^*)] \\
&\leq \frac{M_x^2}{2K^\beta} + \frac{\sum_{k=1}^{K} \frac{L_\theta^2 Q_\theta(\lambda_\theta)}{k}}{2K\delta_K} \qquad (2.36) \\
&\leq \frac{M_x^2}{2K^\beta} + \frac{L_\theta^2 Q_\theta(\lambda_\theta)(\ln(K)+1)}{2K^{1-\beta}}.
\end{aligned}$$

Combining (2.30), (2.31), (2.32), and (2.36), we have that R_K/K can be bounded as follows:

$$\begin{aligned}
\frac{R_K}{K} &\leq \frac{\widehat{R}_K}{K} + \frac{R_K - \widehat{R}_K}{K} \leq \frac{\widehat{R}_K}{K} + \frac{\left|R_K - \widehat{R}_K\right|}{K} \\
&\leq \frac{M_x^2 K^{\alpha-1}}{2} + \frac{M^2(K^{1-\alpha} - \alpha)}{2(1-\alpha)K} + \frac{D_\theta\sqrt{Q_\theta(\lambda_\theta)}(2\sqrt{K}-1)}{K} + \frac{M_x^2}{2K^\beta} + \frac{L_\theta^2 Q_\theta(\lambda_\theta)(\ln(K)+1)}{2K^{1-\beta}}.
\end{aligned}$$

Furthermore, this implies that the limit superior of the average regret is nonpositive. ∎

Remark: In effect, in the context of learning and optimization, the averaging approach leads to a complexity bound given loosely by

$$\mathcal{O}\left(\frac{a}{K^{1-\alpha}} + \frac{b}{K^\alpha} + \frac{d}{K^\beta} + \underbrace{\frac{c_\theta}{\sqrt{K}} + \frac{e_\theta \ln(K)}{K^{1-\beta}}}_{\text{contribution from learning}}\right),$$

where $a, b, c_\theta, d, e_\theta$ are suitably defined. If θ^* is available, then $c_\theta, e_\theta = 0$. Furthermore, by setting $\alpha = 0.5$ and $\beta = 0.5$, this leads to the bound of $\mathcal{O}(\ln K/\sqrt{K})$, which is a degradation as the result of learning θ^*. ∎

2.3 Stochastic variational inequality problems with imperfect information

Several shortcomings exist in the optimization based formulation represented by $(\mathcal{P}_x^o(\theta^*))$. First, the misspecification arises entirely in the objectives while the constraints are known with certainty. Second, the underlying problem need not be an optimization problem, but could instead be captured by a variational inequality problem. Such problems [20] can capture a range of problems including economic equilibrium problems, traffic equilibrium problems, and convex Nash games. In fact, variational inequality problems can effectively capture optimization problems with misspecified constraints. This motivates the consideration of the misspecified stochastic variational inequality problem $(\mathcal{P}_x^v(\theta^*))$ where θ^* can be learnt through the solution of the following problem:

$$(\vartheta - \theta)^T \mathbb{E}[G(\theta;\eta)] \geq 0, \qquad \forall \vartheta \in \Theta, \tag{$\mathcal{L}_\theta^v$}$$

where $G : \theta \times \mathbb{R}^p \to \mathbb{R}^m$, and Θ and η abide by the previous specifications. In the majority of problem settings, $G(\theta;\theta) \triangleq \nabla_\theta g(\theta;\eta)$ but we employ the variational structure to introduce generality. In this section, we extend the results of the previous section to this regime. Specifically, we develop the convergence theory under settings where the variational map F is both strongly monotone and merely monotone in x for every $\theta \in \Theta$ in Section 2.3.1 and provide rate statements in Section 2.3.2.

2.3.1 Almost-sure convergence

As in Section 2.2, we propose a set of coupled stochastic approximation schemes for computing x^* and θ^*. Given $x^0 \in X$ and $\theta^0 \in \Theta$, the coupled SA schemes are stated next:

Algorithm 2 (**Coupled SA schemes for stochastic variational inequality problems**). **Step 0.** *Given $x_0 \in X, \theta_0 \in \Theta$ and sequences $\{\gamma_{k,x}, \gamma_{k,\theta}\}$, $k := 0$*

Step 1.

$$x^{k+1} := \Pi_X \left(x^k - \gamma_{k,x}(F(x^k;\theta^k) + w^k)\right) \tag{Comp$_k$}$$

$$\theta^{k+1} := \Pi_\Theta \left(\theta^k - \gamma_{k,\theta}(G(\theta^k) + v^k)\right), \tag{Learn$_k$}$$

where $w^k \triangleq F(x^k;\theta^k,\xi^k) - F(x^k;\theta^k)$ and $v^k \triangleq G(\theta^k;\eta^k) - G(\theta^k)$.

Step 2. *If $k > K$, stop; else $k : k+1$, go to Step. 1.*

We begin by stating an assumption similar to (A1-1) on the mappings F and G.

Assumption 7 (A1-3)**.** *Suppose the following hold:*

(i) For every $\theta \in \Theta$, $F(x;\theta)$ is both strongly monotone and Lipschitz continuous in x with constants μ_x and L_x, respectively.

(ii) For every $x \in X$, $F(x;\theta)$ is Lipschitz continuous in θ with constant L_θ.

(iii) $G(\theta)$ is strongly monotone and Lipschitz continuous in θ with constants μ_θ and C_θ, respectively.

Now, we can leverage the results in Section 2.2.2 to examine the convergence properties for Algorithm 2.

Proposition 4 (**Almost-sure convergence under strong monotonicity of F**)**.** *Suppose (A1-3), (A2-1) and (A3) hold. Let $\{x^k, \theta^k\}$ be computed via Algorithm 2. Then, $x^k \to x^*$ a.s. and $\theta^k \to \theta^*$ a.s. as $k \to \infty$, where x^* is the unique solution to $(\mathcal{P}_x^v(\theta^*))$ and θ^* is the unique solution to $(\mathcal{L}_\theta^v)$.*

Proof. Note that $x^* = \Pi_X(x^* - \gamma_{k,x}F(x^*;\theta^*))$ and $\theta^* = \Pi_\Theta(\theta^* - \gamma_{k,\theta}G(\theta^*))$. If we replace $\nabla_x f$ and $\nabla_\theta g$ by F and G in Proposition 1, respectively, then by the proof of Proposition 1, we get $x^k \to x^*$ *a.s.* and $\theta^k \to \theta^*$ *a.s.* as $k \to \infty$. ∎

Next, we weaken the rather stringent requirement of strong monotonicity of the map by using an iterative Tikhonov regularization, which can be stated as follows.

Algorithm 3 (**Coupled regularized SA schemes for stochastic variational inequality problems**)**.**

Step 0. *Given $x_0 \in X, \theta_0 \in \Theta$ and sequences $\{\gamma_{k,x}, \gamma_{k,\theta}\}$, $k := 0$*

Step 1.

$$x^{k+1} := \Pi_X\left(x^k - \gamma_{k,x}(F(x^k;\theta^k) + \epsilon_k x^k + w^k)\right) \qquad (\text{Comp}_k)$$

$$\theta^{k+1} := \Pi_\Theta\left(\theta^k - \gamma_{k,\theta}(G(\theta^k) + v^k)\right), \qquad (\text{Learn}_k)$$

where $w^k \triangleq F(x^k;\theta^k,\xi^k) - F(x^k;\theta^k)$ and $v^k \triangleq G(\theta^k;\eta^k) - G(\theta^k)$.

Step 2. *If $k > K$, stop; else $k := k+1$, go to Step. 1.*

Unlike in standard Tikhonov regularization, such a scheme updates the regularization parameter ϵ_k after every step. Tikhonov regularization and its iterative counterpart has a long history [39] while iterative regularization schemes have seen relatively less study in the context of variational inequality problems (cf. [48, 49]).

Of note is the extension to distributed schemes to accommodate monotone Cartesian stochastic variational inequality problems [22]. We employ such techniques in developing single-loop stochastic approximation schemes in the context of learning and optimization. The following assumptions will be made on both the decision variable and parameter.

Assumption 8 (A1-4)**.** *Suppose the following holds in addition to (A1-3 (ii)) and (A1-3 (iii)).*

(i) For every $\theta \in \Theta$, $F(x;\theta)$ is monotone in x and Lipschitz continuous in x with constant L_x.

In iterative Tikhonov regularization, one cannot independently choose $\{\epsilon_k\}$ and $\{\gamma_k\}$; in fact, these sequences are related and satisfy some collectively imposed requirements.

Assumption 9 (A2-3)**.** *Let $\{\gamma_{k,x}\}$, $\{\gamma_{k,\theta}\}$, $\{\epsilon_k\}$ and some constant $\tau \in (0,1)$ be chosen such that:*

(i) $\sum_{k=0}^{\infty} \gamma_{k,x}^{2-\tau} < \infty$ *and* $\sum_{k=0}^{\infty} \gamma_{k,\theta}^{2} < \infty$,

(ii) $\sum_{k=0}^{\infty} \gamma_{k,x}\epsilon_k = \infty$ *and* $\sum_{k=0}^{\infty} \gamma_{k,\theta} = \infty$,

(iii) $\beta_k = \frac{\gamma_{k,x}^{\tau}}{2\gamma_{k,\theta}\mu_\theta} \downarrow 0$ *as* $k \to 0$.

(iv) $\sum_{k=0}^{\infty} \frac{(\epsilon_{k-1}-\epsilon_k)}{\epsilon_k} < \infty$.

Before providing a convergence result for Algorithm 3, we introduce the following results.

Lemma 5. *Let $H : K \to \mathbb{R}^n$ be a mapping that is monotone over K, and Lipschitz continuous over K with constant L. Then, for any $\gamma > 0$ and $\epsilon > 0$, we have $\|(x-y) - \gamma(H(x)-H(y)) - \epsilon\gamma(x-y)\| \leq q\|x-y\|$, where $q = \sqrt{1-2\gamma\epsilon+\gamma^2(L^2+\epsilon^2)}$.*

Proof. See proof of Theorem 1 in [50]. ∎

Lemma 6. *Let $H : K \to \mathbb{R}^n$ be a mapping that is monotone over K. Given $\epsilon_k > 0$, let y^k be a solution to VI$(K, H+\epsilon_k\mathbf{I})$. Then,*

$$\|y^k - y^{k-1}\| \leq \frac{M(\epsilon_{k-1}-\epsilon_k)}{\epsilon_k},$$

where $M = \|x^*\|$ and x^* is a solution to VI(H,K).

Proof. See Lemma 3 in [50]. ∎

The convergence result for Algorithm 3 can be stated as follows.

Theorem 5 (**Almost-sure convergence under monotonicity of F**). *Suppose (A1-4) , (A2-3) and (A3) hold. Suppose X is bounded and the solution set X^* of $(\mathcal{P}_x^v(\theta^*))$ is nonempty. Let $\{x^k, \theta^k\}$ be computed via Algorithm 3. Then, $\theta^k \to \theta^*$ a.s. as $k \to \infty$, and x^k converges to a random point in X^* a.s. as $k \to \infty$.*

Proof. We have for any $x^* \in X^*$ that $x^* = \Pi_X(x^* - \gamma_{k,x} F(x^*; \theta^*))$. Suppose y^k is a solution to the following fixed-point problem

$$y^k = \Pi_X(y^k - \gamma_{k,x}(F(y^k; \theta^*) + \epsilon_k y^k)).$$

Then, by the triangle inequality $\|x^{k+1} - x^*\|$ may be bounded as follows:

$$\|x^{k+1} - x^*\| \leq \underbrace{\|x^{k+1} - y^k\|}_{\textbf{Term 1}} + \underbrace{\|y^k - x^*\|}_{\textbf{Term 2}}.$$

Term 2 converges to zero by the convergence statement of Tikhonov regularization methods [20]. By using the non-expansivity of the Euclidean projector, $\|x^{k+1} - y^k\|^2$ can be bounded as follows:

$$\begin{aligned}\|x^{k+1} - y^k\|^2 &= \|\Pi_X(x^k - \gamma_{k,x}(F(x^k; \theta^k) + \epsilon_k x^k + w^k)) - \Pi_X(y^k - \gamma_{k,x}(F(y^k; \theta^*) + \epsilon_k y^k))\|^2 \\ &\leq \|(x^k - y^k) - \gamma_{k,x}(F(x^k; \theta^k) - F(y^k; \theta^*)) - \epsilon_k \gamma_{k,x}(x^k - y^k) - \gamma_{k,x} w^k\|^2.\end{aligned}$$

By adding and subtracting $\gamma_{k,x} F(x^k; \theta^*)$, this expression can be further expanded as follows:

$$\begin{aligned}&\|(x^k - y^k) - \gamma_{k,x}(F(x^k; \theta^*) - F(y^k; \theta^*)) - \gamma_{k,x}(F(x^k; \theta^k) - F(x^k; \theta^*)) - \epsilon_k \gamma_{k,x}(x^k - y^k) - \gamma_{k,x} w^k\|^2 \\ &= \|(x^k - y^k) - \gamma_{k,x}(F(x^k; \theta^*) - F(y^k; \theta^*)) - \epsilon_k \gamma_{k,x}(x^k - y^k)\|^2 + \gamma_{k,x}^2 \|F(x^k; \theta^k) - F(x^k; \theta^*)\|^2 + \gamma_{k,x}^2 \|w^k\|^2 \\ &- 2[(x^k - y^k) - \gamma_{k,x}(F(x^k; \theta^*) - F(y^k; \theta^*)) - \epsilon_k \gamma_{k,x}(x^k - y^k)]^T \times (F(x^k; \theta^k) - F(x^k; \theta^*)) \\ &- 2[(x^k - y^k) - \gamma_{k,x}(F(x^k; \theta^*) - F(y^k; \theta^*)) - \epsilon_k \gamma_{k,x}(x^k - y^k)]^T w^k + 2\gamma_{k,x}^2 (F(x^k; \theta^k) - F(x^k; \theta^*))^T w^k.\end{aligned}$$

Noting that $\mathbb{E}[w^k \mid \mathcal{F}_k] = 0$, we have

$$\mathbb{E}[\|x^{k+1} - y^k\|^2 \mid \mathcal{F}_k] \leq \textbf{Term 3} + \textbf{Term 4} + \textbf{Term 5} + \gamma_{k,x}^2 \mathbb{E}[\|w^k\|^2 \mid \mathcal{F}_k], \tag{2.37}$$

where

$$\textbf{Term 3} \triangleq \|(x^k - y^k) - \gamma_{k,x}(F(x^k;\theta^*) - F(y^k;\theta^*)) - \epsilon_k\gamma_{k,x}(x^k - y^k)\|^2,$$

$$\textbf{Term 4} \triangleq \gamma_{k,x}^2\|F(x^k;\theta^k) - F(x^k;\theta^*)\|^2,$$

$$\textbf{Term 5} \triangleq -2\gamma_{k,x}[(x^k - y^k) - \gamma_{k,x}(F(x^k;\theta^*) - F(y^k;\theta^*)) - \epsilon_k\gamma_{k,x}(x^k - y^k)]^T(F(x^k;\theta^k) - F(x^k;\theta^*)).$$

By Lemma 5 and (A1-4), Term 3 can be further bounded by

$$(1 - 2\gamma_{k,x}\epsilon_k + \gamma_{k,x}^2(L_x^2 + (\epsilon_k)^2))\|x^k - y^k\|^2. \tag{2.38}$$

By the Lipschitz continuity of $F(x;\theta)$ in θ (A1-4), Term 4 can be further bounded by

$$\gamma_{k,x}^2 L_\theta^2\|\theta^k - \theta^*\|^2. \tag{2.39}$$

By the Cauchy-Schwarz inequality, Lemma 5, (A1-4) as well as the fact that $2ab \leq a^2 + b^2$, Term 5 can be further bounded by

$$\begin{aligned}
&2\gamma_{k,x}\|(x^k - y^k) - \gamma_{k,x}(F(x^k;\theta^*) - F(y^k;\theta^*)) - \epsilon_k\gamma_{k,x}(x^k - y^k)\|\|F(x^k;\theta^k) - F(x^k;\theta^*)\| \\
&\leq 2\gamma_{k,x}\sqrt{1 - 2\gamma_{k,x}\epsilon_k + \gamma_{k,x}^2(L_x^2 + (\epsilon_k)^2)}\|x^k - y^k\|L_\theta\|\theta^k - \theta^*\| \\
&\leq 2\gamma_{k,x}L_\theta\|x^k - y^k\|\|\theta^k - \theta^*\| \\
&\leq \gamma_{k,x}^{2-\tau}L_\theta^2\|x^k - y^k\|^2 + \gamma_{k,x}^\tau\|\theta^k - \theta^*\|^2,
\end{aligned} \tag{2.40}$$

where $\tau \in (0,1)$ is chosen to satisfy (A2-3). Combining (2.37), (2.38), (2.39) and (2.40), we get

$$\mathbb{E}[\|x^{k+1} - y^k\|^2 \mid \mathcal{F}_k] \leq (q_{k,x}^2 + \gamma_{k,x}^{2-\tau}L_\theta^2)\|x^k - y^k\|^2 + (\gamma_{k,x}^\tau + \gamma_{k,x}^2 L_\theta^2)\|\theta^k - \theta^*\|^2 + \gamma_{k,x}^2\nu_x^2, \tag{2.41}$$

where $q_{k,x} = \sqrt{1 - 2\gamma_{k,x}\epsilon_k + \gamma_{k,x}^2(L_x^2 + (\epsilon_k)^2)}$.

On the other hand, we have that θ^* is the unique solution to $\mathrm{VI}(\Theta, \mathbb{E}[G(\bullet;\eta)])$ and

$$\theta^* = \Pi_\Theta(\theta^* - \gamma_{k,\theta}G(\theta^*)).$$

Therefore, by the nonexpansivity of the Euclidean projector, $\|\theta^{k+1} - \theta^*\|^2$ may be bounded as follows:

$$\begin{aligned}
&\|\theta^{k+1} - \theta^*\|^2 = \|\Pi_\Theta(\theta^k - \gamma_{k,\theta}(G(\theta^k) + v^k)) - \Pi_\Theta(\theta^* - \gamma_{k,\theta}G(\theta^*))\|^2 \\
&\leq \|(\theta^k - \theta^*) - \gamma_{k,\theta}(G(\theta^k) - G(\theta^*)) - \gamma_{k,\theta}v^k\|^2 \\
&= \|(\theta^k - \theta^*) - \gamma_{k,\theta}(G(\theta^k) - G(\theta^*))\|^2 + \gamma_{k,\theta}^2\|v^k\|^2 - 2\gamma_{k,\theta}[(\theta^k - \theta^*) - \gamma_{k,\theta}(G(\theta^k) - G(\theta^*))]^T v^k.
\end{aligned}$$

By taking conditional expectations and by recalling that $\mathbb{E}[v^k \mid \mathcal{F}_k] = 0$ (A3), we obtain the following:

$$\begin{aligned}
\mathbb{E}[\|\theta^{k+1} - \theta^*\|^2 \mid \mathcal{F}_k] &\leq \|(\theta^k - \theta^*) - \gamma_{k,\theta}(G(\theta^k) - G(\theta^*))\|^2 + \gamma_{k,\theta}^2\mathbb{E}[\|v^k\|^2 \mid \mathcal{F}_k] \\
&\leq q_{k,\theta}^2\|\theta^k - \theta^*\|^2 + \gamma_{k,\theta}^2\nu_\theta^2,
\end{aligned} \tag{2.42}$$

where $q_{k,\theta} = \sqrt{1 - 2\gamma_{k,\theta}\mu_\theta + \gamma_{k,\theta}^2 C_\theta^2}$, and the second inequality follows from Lemma 4, (A1-4) and (A3). Since by (A2-3) $\sum_{k=0}^{\infty}(1 - q_{k,\theta}^2) = \infty$ and $\sum_{k=0}^{\infty}\gamma_{k,\theta}^2\nu_\theta^2 < \infty$, and

$$\lim_{k\to\infty}\frac{\gamma_{k,\theta}^2\nu_\theta^2}{1 - q_{k,\theta}^2} = \lim_{k\to\infty}\frac{\gamma_{k,\theta}^2\nu_\theta^2}{2\gamma_{k,\theta}\mu_\theta - \gamma_{k,\theta}^2 C_\theta^2} = \lim_{k\to\infty}\frac{\gamma_{k,\theta}\nu_\theta^2}{2\mu_\theta - \gamma_{k,\theta}C_\theta^2} = 0,$$

we have by Lemma 2 that $\|\theta^k - \theta^*\| \to 0$ *a.s.* as $k \to \infty$. Choose $\beta_k = \frac{\gamma_{k,x}^\tau}{2\gamma_{k,\theta}\mu_\theta}$ by (A2-3). Note that by assumption $\beta_{k+1} \leq \beta_k$. By multiplying the left hand side of (2.42) by β_{k+1} and adding to the left hand side of (2.41), we get

$$\begin{aligned}
&\mathbb{E}[\|x^{k+1} - y^k\|^2 \mid \mathcal{F}_k] + \beta_{k+1}\mathbb{E}[\|\theta^{k+1} - \theta^*\|^2 \mid \mathcal{F}_k] \\
&\leq \mathbb{E}[\|x^{k+1} - y^k\|^2 \mid \mathcal{F}_k] + \beta_k\mathbb{E}[\|\theta^{k+1} - \theta^*\|^2 \mid \mathcal{F}_k] \\
&\leq (q_{k,x}^2 + \gamma_{k,x}^{2-\tau}L_\theta^2)\|x^k - y^k\|^2 + (\beta_k q_{k,\theta}^2 + \gamma_{k,x}^\tau + \gamma_{k,x}^2 L_\theta^2)\|\theta^k - \theta^*\|^2 + \beta_k\gamma_{k,\theta}^2\nu_\theta^2 + \gamma_{k,x}^2\nu_x^2 \\
&= (q_{k,x}^2 + \gamma_{k,x}^{2-\tau}L_\theta^2)\|x^k - y^k\|^2 + \underbrace{\frac{\beta_k q_{k,\theta}^2 + \gamma_{k,x}^\tau + \gamma_{k,x}^2 L_\theta^2}{\beta_k}}_{\textbf{Term6}}\cdot\beta_k\|\theta^k - \theta^*\|^2 + \beta_k\gamma_{k,\theta}^2\nu_\theta^2 + \gamma_{k,x}^2\nu_x^2.
\end{aligned} \tag{2.43}$$

Term 6 on the right hand side of (2.43) can be further expanded as

$$\begin{aligned}
\frac{\beta_k q_{k,\theta}^2 + \gamma_{k,x}^\tau + \gamma_{k,x}^2 L_\theta^2}{\beta_k} &= q_{k,\theta}^2 + \frac{\gamma_{k,x}^\tau + \gamma_{k,x}^2 L_\theta^2}{\beta_k} = 1 - 2\gamma_{k,\theta}\mu_\theta + \gamma_{k,\theta}^2 C_\theta^2 + \frac{\gamma_{k,x}^\tau}{\beta_k} + \frac{\gamma_{k,x}^2 L_\theta^2}{\beta_k} \\
&= 1 + \gamma_{k,\theta}^2 C_\theta^2 + 2\gamma_{k,\theta}\gamma_{k,x}^{2-\tau}\mu_\theta L_\theta^2.
\end{aligned} \tag{2.44}$$

Combining (2.43) and (2.44), we get

$$
\begin{aligned}
&\mathbb{E}[\|x^{k+1}-y^k\|^2 \mid \mathcal{F}_k] + \beta_{k+1}\mathbb{E}[\|\theta^{k+1}-\theta^*\|^2 \mid \mathcal{F}_k] \\
&\leq (q_{k,x}^2 + \gamma_{k,x}^{2-\tau}L_\theta^2)\|x^k-y^k\|^2 + (1+\gamma_{k,\theta}^2 C_\theta^2 + 2\gamma_{k,\theta}\gamma_{k,x}^{2-\tau}\mu_\theta L_\theta^2)\beta_k\|\theta^k-\theta^*\|^2 + \beta_k\gamma_{k,\theta}^2\nu_\theta^2 + \gamma_{k,x}^2\nu_x^2 \\
&= (1+\gamma_{k,\theta}^2 C_\theta^2 + 2\gamma_{k,\theta}\gamma_{k,x}^{2-\tau}\mu_\theta L_\theta^2)(\|x^k-y^k\|^2 + \beta_k\|\theta^k-\theta^*\|^2) \\
&- (\gamma_{k,\theta}^2 C_\theta^2 + 2\gamma_{k,\theta}\gamma_{k,x}^{2-\tau}\mu_\theta L_\theta^2 + 2\gamma_{k,x}\epsilon_k)\|x^k-y^k\|^2 \\
&+ (\gamma_{k,x}^2(L_x^2+(\epsilon_k)^2) + \gamma_{k,x}^{2-\tau}L_\theta^2)\|x^k-y^k\|^2 + \beta_k\gamma_{k,\theta}^2\nu_\theta^2 + \gamma_{k,x}^2\nu_x^2.
\end{aligned}
$$

Note that $\|x^{k+1}-y^k\|^2 \leq \|x^k-y^{k-1}\|^2 + 2\|x^k-y^{k-1}\|\|y^k-y^{k-1}\| + \|y^k-y^{k-1}\|^2$. We have

$$
\begin{aligned}
&\mathbb{E}[\|x^{k+1}-y^k\|^2 \mid \mathcal{F}_k] + \beta_{k+1}\mathbb{E}[\|\theta^{k+1}-\theta^*\|^2 \mid \mathcal{F}_k] \\
&\leq (1+\gamma_{k,\theta}^2 C_\theta^2 + 2\gamma_{k,\theta}\gamma_{k,x}^{2-\tau}\mu_\theta L_\theta^2)(\|x^k-y^{k-1}\|^2 + \beta_k\|\theta^k-\theta^*\|^2) \\
&\quad + 2(1+\gamma_{k,\theta}^2 C_\theta^2 + 2\gamma_{k,\theta}\gamma_{k,x}^{2-\tau}\mu_\theta L_\theta^2)\|x^k-y^{k-1}\|\|y^k-y^{k-1}\| + (1+\gamma_{k,\theta}^2 C_\theta^2 + 2\gamma_{k,\theta}\gamma_{k,x}^{2-\tau}\mu_\theta L_\theta^2)\|y^k-y^{k-1}\|^2 \\
&\quad - (\gamma_{k,\theta}^2 C_\theta^2 + 2\gamma_{k,\theta}\gamma_{k,x}^{2-\tau}\mu_\theta L_\theta^2 + 2\gamma_{k,x}\epsilon_k)\|x^k-y^k\|^2 + (\gamma_{k,x}^2(L_x^2+(\epsilon_k)^2) + \gamma_{k,x}^{2-\tau}L_\theta^2)\|x^k-y^k\|^2 \\
&\quad + \beta_k\gamma_{k,\theta}^2\nu_\theta^2 + \gamma_{k,x}^2\nu_x^2,
\end{aligned}
$$

which can be further reduced to

$$
\begin{aligned}
&\mathbb{E}[\|x^{k+1}-y^k\|^2 \mid \mathcal{F}_k] + \beta_{k+1}\mathbb{E}[\|\theta^{k+1}-\theta^*\|^2 \mid \mathcal{F}_k] \\
&\leq (1+\gamma_{k,\theta}^2 C_\theta^2 + 2\gamma_{k,\theta}\gamma_{k,x}^{2-\tau}\mu_\theta L_\theta^2)(\|x^k-y^{k-1}\|^2 + \beta_k\|\theta^k-\theta^*\|^2) \\
&\quad + 2(1+\gamma_{k,\theta}^2 C_\theta^2 + 2\gamma_{k,\theta}\gamma_{k,x}^{2-\tau}\mu_\theta L_\theta^2)\|x^k-y^{k-1}\|\|y^k-y^{k-1}\| \\
&\quad + (1+\gamma_{k,\theta}^2 C_\theta^2 + 2\gamma_{k,\theta}\gamma_{k,x}^{2-\tau}\mu_\theta L_\theta^2)\|y^k-y^{k-1}\|^2 \\
&\quad - 2\gamma_{k,x}\epsilon_k\|x^k-y^k\|^2 + (\gamma_{k,x}^2(L_x^2+(\epsilon_k)^2) + \gamma_{k,x}^{2-\tau}L_\theta^2)\|x^k-y^k\|^2 + \beta_k\gamma_{k,\theta}^2\nu_\theta^2 + \gamma_{k,x}^2\nu_x^2.
\end{aligned}
$$

By Lemma 6 and (A2-3), $\sum_{k=0}^{\infty}\|y^k-y^{k-1}\| < \infty$. and $\sum_{k=0}^{\infty}\|y^k-y^{k-1}\|^2 < \infty$. Therefore, by boundedness of X, (A2-3) and Lemma 3, we have that there exists a random variable V such that

$$
\|x^k-y^{k-1}\|^2 + \beta_k\|\theta^k-\theta^*\|^2 \to V \quad a.s. \quad \text{as } k\to\infty.
$$

and $\sum_{k=0}^{\infty} 2\gamma_{k,x}\epsilon_k\|x^k-y^k\|^2 < \infty$. Since $\sum_{k=0}^{\infty}\gamma_{k,x}\epsilon_k = \infty$, we get $\|x^k-y^k\| \to 0$ $a.s.$ as $k\to\infty$. This implies $\|x^k-x^*\| \to 0$ $a.s.$ as $k\to\infty$. ∎

2.3.2 Diminishing and constant steplength error analysis

In this section, we estimate the convergence rate of the proposed schemes. Analogous to Section 2.2.3, we obtain the optimal $\mathcal{O}(1/K)$ rate estimate for the upper bound on the expected error in the solution x_K when $F(\bullet;\theta^*)$ is strongly monotone in $(\bullet)$. In addition, when $F(\bullet;\theta^*)$ is merely monotone and the variational inequality problem possesses the *minimum principle sufficiency* (MPS) property (See Lemma 7 for a definition of the MPS property), a rate estimate is still available by using averaging. If we replace $\nabla_x f$ and $\nabla_\theta g$ by F and G, respectively, in Theorem 2, then we obtain the following:

Theorem 6 (**Rate estimate for strongly monotone** F)**.** *Suppose (A1-3) and (A3) hold. Suppose* $\gamma_{x,k} = \lambda_x/k$ *and* $\gamma_{\theta,k} = \lambda_\theta/k$ *with* $\lambda_x > 1/\mu_x$ *and* $\lambda_\theta > 1/(2\mu_\theta)$*. Let* $\mathbb{E}[\|F(x^k;\theta^k)+w^k\|^2] \leq M^2$ *and* $\mathbb{E}[\|G(\theta^k)+v^k\|^2] \leq M_\theta^2$ *for all* $x^k \in X$ *and* $\theta^k \in \Theta$*. Suppose* x^* *is the unique solution to VI*$(X, \mathbb{E}[F(\bullet;\theta^*,\xi)])$*. Let* $\{x^k,\theta^k\}$ *be computed via Algorithm 2. Then, the following hold:*

$$\mathbb{E}[\|\theta^k-\theta^*\|^2] \leq \frac{Q_\theta(\lambda_\theta)}{k} \text{ and } \mathbb{E}[\|x^k-x^*\|^2] \leq \frac{Q_x(\lambda_x)}{k},$$

where

$$\begin{aligned}
Q_\theta(\lambda_\theta) &\triangleq \max\left\{\lambda_\theta^2 M_\theta^2(2\mu_\theta\lambda_\theta-1)^{-1}, \mathbb{E}[\|\theta^1-\theta^*\|^2]\right\},\\
Q_x(\lambda_x) &\triangleq \max\left\{\lambda_x^2\widetilde{M}^2(\mu_x\lambda_x-1)^{-1}, \mathbb{E}[\|x^1-x^*\|^2]\right\},\\
\text{and } \widetilde{M} &\triangleq \sqrt{M^2+\frac{L_\theta^2 Q_\theta(\lambda_\theta)}{\mu_x\lambda_x}}.
\end{aligned}$$

Next, we weaken the strong monotonicity of F, but assume that $(\mathcal{P}_x^v(\theta^*))$ satisfies the MPS property, introduced in the following Lemma. Note that this property guarantees weak sharpness of the solution set; this is analogous to weak-sharpness of minima in optimization problems [51].

Lemma 7 (Theorem 4.3 in [52])**.** *Let* $H: X \to \mathbb{R}^n$ *be a mapping that is monotone over the compact polyhedral set* X*. Let* X^* *be the solution set of VI*(X,H)*. If the VI*(X,H) *possesses the minimum principle sufficiency (MPS) property, then there exists a positive number* α *such that* $(x-x^*)^T H(x^*) \geq \alpha\ \mathrm{dist}(x,X^*), \quad \forall x \in X, \quad \forall x^* \in X^*$*, where* $\mathrm{dist}(x,X^*) \triangleq \min_{x^*\in X^*}\|x-x^*\|$*. We say that the VI*$(X,H)$ *possesses the MPS property if* $\Gamma(x^*) = X^*$ *for every* x^* *in* X^**, where* $\Gamma(x) = \arg\max_{y\in X}(x-y)^T H(x)$.

By leveraging this property, we may estimate the convergence rate by using averaging as in Theorem 2.

Theorem 7 (**Rate estimates under monotonicity of F**). *Suppose (A1-4) and (A3) hold. Suppose* $\mathbb{E}[\|x^k - x^*\|^2] \le M_x^2$, $\mathbb{E}[\|F(x^k;\theta^k) + w^k\|^2] \le M^2$ *and* $\mathbb{E}[\|G(\theta^k) + v^k\|^2] \le M_\theta^2$ *for all* $x^k \in X$ *and* $\theta^k \in \Theta$. *Suppose* X *is a compact polyhedral set, the solution set* X^* *of VI*$(X, \mathbb{E}[F(\bullet;\theta^*,\xi)])$ *is nonempty, and* x^* *is a point in* X^*. *Suppose VI*$(X, \mathbb{E}[F(\bullet;\theta^*,\xi)])$ *possesses the MPS property. Let* $\{x^k, \theta^k\}$ *be computed via Algorithm 2. For* $1 \le i, t \le k$, *we define* $v_t \triangleq \frac{\gamma_{x,t}}{\sum_{s=i}^k \gamma_{x,s}}$, $\tilde{x}_{i,k} \triangleq \sum_{t=i}^k v_t x^t$ *and* $D_X \triangleq \max_{x\in X} \|x - x^1\|$. *Suppose for* $1 \le t \le k$

$$\gamma_x = \sqrt{\frac{4D_X^2 + L_\theta^2 Q_\theta(\lambda_\theta)(1+\ln k)}{(M^2 + M_x^2)k}},$$

where $Q_\theta(\lambda_\theta) \triangleq \max\left\{\lambda_\theta^2 M_\theta^2 (2\mu_\theta\lambda_\theta - 1)^{-1}, \mathbb{E}[\|\theta^1 - \theta^*\|^2]\right\}$, *and* $\gamma_{\theta,k} = \lambda_\theta/k$ *with* $\lambda_\theta > 1/(2\mu_\theta)$. *Then there exists a positive number* α *such that for* $1 \le i \le k$:

$$\mathbb{E}\left[\alpha \operatorname{dist}(\tilde{x}_{i,k}, X^*)\right] \le C_{i,k}\sqrt{\frac{B_k}{k}},$$

where $C_{i,k} = \frac{k}{k-i+1}$ *and* $B_k = (4D_X^2 + L_\theta^2 Q_\theta(\lambda_\theta)(1+\ln k))(M^2 + M_x^2)$.

Proof. By using the same notation in Theorem 2 except that we replace $\nabla_x f$ and $\nabla_\theta g$ by F and G, respectively, we have from (2.20) that

$$a_{k+1} \le a_k + \frac{1}{2}\gamma_{x,k}^2 M^2 - \gamma_{x,k}\mathbb{E}[(x^k - x^*)^T F(x^k;\theta^*)] - \gamma_{x,k}\mathbb{E}[(x^k - x^*)^T(F(x^k;\theta^k) - F(x^k;\theta^*))]. \tag{2.45}$$

By Lemma 7, we have that there exists a positive number α such that

$$\begin{aligned}\alpha \operatorname{dist}(x^k, X^*) &\le (x^k - x^*)^T F(x^*;\theta^*) = (x^k - x^*)^T F(x^k;\theta^*) - (x^k - x^*)^T(F(x^k;\theta^*) - F(x^*;\theta^*)) \\ &\le (x^k - x^*)^T F(x^k;\theta^*),\end{aligned} \tag{2.46}$$

where the last inequality follows from the monotonicity of $F(\bullet;\theta^*)$ in $(\bullet)$. Combining (2.45) and (2.46),

$$\begin{aligned}\alpha\gamma_{x,k}\mathbb{E}[\operatorname{dist}(x^k, X^*)] &\le \gamma_{x,k}\mathbb{E}[(x^k - x^*)^T F(x^k;\theta^*)] \\ &\le a_k - a_{k+1} + \frac{1}{2}\gamma_{x,k}^2 M^2 - \gamma_{x,k}\mathbb{E}[(x^k - x^*)^T(F(x^k;\theta^k) - F(x^k;\theta^*))].\end{aligned} \tag{2.47}$$

Next, we follow the same proof method in Theorem 2. We define $v_t \triangleq \frac{\gamma_{x,t}}{\sum_{s=i}^k \gamma_{x,s}}$ and $D_X \triangleq \max_{x\in X} \|x - x^1\|$. It follows from (2.24) and (2.47) that

$$\mathbb{E}\left[\alpha \sum_{t=i}^k v_t \operatorname{dist}(x^t, X^*)\right] \le \frac{a_i + \frac{1}{2}\sum_{t=i}^k \gamma_{x,t}^2 (M^2 + M_x^2) + \frac{1}{2}L_\theta^2 Q_\theta(\lambda_\theta)(1+\ln k)}{\sum_{t=i}^k \gamma_{x,t}}. \tag{2.48}$$

Next, we consider points given by $\tilde{x}_{i,k} \triangleq \sum_{t=i}^{k} v_t x^t$. Since $F(x;\theta^*)$ is monotone in x, we have that X^* is convex, which implies that $\text{dist}(x, X^*)$ is convex in x. So, we get $\text{dist}(\tilde{x}_{i,k}, X^*) \leq \sum_{t=i}^{k} v_t \text{dist}(x^t, X^*)$. It follows from (2.25) and (2.48) that for $1 \leq i \leq k$

$$\mathbb{E}\left[\alpha \text{ dist}(\tilde{x}_{i,k}, X^*)\right] \leq \frac{4D_X^2 + \sum_{t=i}^{k} \gamma_{x,t}^2 (M^2 + M_x^2) + L_\theta^2 Q_\theta(\lambda_\theta)(1 + \ln k)}{2\sum_{t=i}^{k} \gamma_{x,t}}. \tag{2.49}$$

Suppose $\gamma_{x,t} = \gamma_x$ for $t = 1, \ldots, k$. If we follow the same proof method in Theorem 2, then we can get from (2.27) and (2.49) that

$$\mathbb{E}\left[\alpha \text{ dist}(\tilde{x}_{i,k}, X^*)\right] \leq C_{i,k}\sqrt{\frac{B_k}{k}}.$$

∎

The following corollary is a special case of Theorem 7, an avenue that has been adopted in [41].

Corollary 2 (Rate estimates under monotonicity of F). *Suppose (A1-4) and (A3) hold. Suppose $\mathbb{E}[\|x^k - x^*\|^2] \leq M_x^2$, $\mathbb{E}[\|F(x^k;\theta^k) + w^k\|^2] \leq M^2$ and $\mathbb{E}[\|G(\theta^k) + v^k\|^2] \leq M_\theta^2$ for all $x^k \in X$ and $\theta^k \in \Theta$. Suppose X is a compact polyhedral set, the solution set X^* of VI$(X, \mathbb{E}[F(\bullet;\theta^*,\xi)])$ is nonempty, and x^* is a point in X^*. Suppose VI$(X, \mathbb{E}[F(\bullet;\theta^*,\xi)])$ possesses the MPS property. Let $\{x^k, \theta^k\}$ be computed via Algorithm 2. For $k/2 \leq t \leq k$, we define $v_t \triangleq \frac{\gamma_{x,t}}{\sum_{s=k/2}^{k} \gamma_{x,s}}$, $\tilde{x}_{k/2,k} \triangleq \sum_{t=k/2}^{k} v_t x^t$ and $D_X \triangleq \max_{x \in X} \|x - x^1\|$. Suppose for $1 \leq t \leq k$*

$$\gamma_x = \sqrt{\frac{4D_X^2 + L_\theta^2 Q_\theta(\lambda_\theta)(1 + \ln 2)}{(M^2 + M_x^2)k}},$$

where $Q_\theta(\lambda_\theta) \triangleq \max\left\{\lambda_\theta^2 M_\theta^2 (2\mu_\theta \lambda_\theta - 1)^{-1}, \mathbb{E}[\|\theta^1 - \theta^*\|^2]\right\}$, and $\gamma_{\theta,k} = \lambda_\theta / k$ with $\lambda_\theta > 1/(2\mu_\theta)$. Then there exists a positive number α such that

$$\mathbb{E}\left[\alpha \text{ dist}(\tilde{x}_{k/2,k}, X^*)\right] \leq 2\sqrt{\frac{B}{k}},$$

where $B = (4D_X^2 + L_\theta^2 Q_\theta(\lambda_\theta)(1 + \ln 2))(M^2 + M_x^2)$.

Proof. When $i = k/2$ where k is a positive even number, then by utilizing the same approach as in Corollary 1, inequality (2.49) becomes the following:

$$\mathbb{E}\left[\alpha \text{ dist}(\tilde{x}_{k/2,k}, X^*)\right] \leq \frac{4D_X^2 + \sum_{t=k/2}^{k} \gamma_{x,t}^2 (M^2 + M_x^2) + L_\theta^2 Q_\theta(\lambda_\theta)(1 + \ln 2)}{2\sum_{t=k/2}^{k} \gamma_{x,t}}. \tag{2.50}$$

Suppose $\gamma_{x,t} = \gamma_x$ for $t = 1, \ldots, k$. By utilizing the same techniques as in Theorem 7, then we obtain the following bound:

$$\mathbb{E}\left[\alpha \operatorname{dist}(\tilde{x}_{k/2,k}, X^*)\right] \leq 2\sqrt{\frac{B}{k}},$$

where $B \triangleq (4D_X^2 + L_\theta^2 Q_\theta(\lambda_\theta)(1 + \ln 2))(M^2 + M_x^2)$. ∎

Next, we present a constant steplength error bound.

Proposition 5 (Constant steplength error bound). *Suppose (A3) holds. Suppose $\gamma_{\theta,k} := \gamma_\theta$ and $\gamma_{x,k} := \gamma_x$. Suppose $\mathbb{E}[\|x^k - x^*\|^2] \leq M_x^2$ and $\mathbb{E}[F(x^k; \theta^k) + w^k\|^2] \leq M^2$ for all $x^k \in X$. Suppose $A_k \triangleq \frac{1}{2}\|x^k - x^*\|^2$ and $a_k \triangleq \mathbb{E}[A_k]$. Suppose X is a compact polyhedral set, the solution set X^* of VI$(X, \mathbb{E}[F(\bullet; \theta^*, \xi)])$ is nonempty, and x^* is a point in X^*. Suppose VI$(X, \mathbb{E}[F(\bullet; \theta^*, \xi)])$ possesses the MPS property. Let $\{x^k, \theta^k\}$ be computed via Algorithm 1.*

(i) Suppose (A1-3) holds. Then, the following holds:

$$\limsup_{k\to\infty} a_k \leq \frac{1}{2\mu_x}\gamma_x M^2 + \frac{1}{2}\frac{1}{\mu_x^2} L_\theta^2 \frac{\gamma_\theta \nu_\theta^2}{2\mu_\theta - \gamma_\theta C_\theta^2};$$

(ii) Suppose (A1-4) holds. Then, there exists a positive number α such that:

$$\limsup_{k\to\infty} \mathbb{E}[\operatorname{dist}(x^k, X^*)] \leq \frac{1}{\alpha}\left[\frac{1}{2}\gamma_x M^2 + \frac{1}{2}\gamma_x^{1-\tau} M_x^2 + \frac{1}{2}\gamma_x^{\tau-1} L_\theta^2 \frac{\gamma_\theta \nu_\theta^2}{2\mu_\theta - \gamma_\theta C_\theta^2}\right],$$

where $0 < \tau < 1$.

Proof. If we replace $\nabla_x f$ and $\nabla_\theta g$ by F and G in Proposition 3, we obtain that

$$\limsup_{k\to\infty} \mathbb{E}[\|\theta^k - \theta^*\|^2] \leq \frac{\gamma_\theta \nu_\theta^2}{2\mu_\theta - \gamma_\theta C_\theta^2},$$

and the following can be derived based on the properties of F:

(i) F is strongly monotone:

$$\limsup_{k\to\infty} a_k \leq \frac{1}{2\mu_x}\gamma_x M^2 + \frac{1}{2}\frac{1}{\mu_x^2} L_\theta^2 \frac{\gamma_\theta \nu_\theta^2}{2\mu_\theta - \gamma_\theta C_\theta^2};$$

(ii) f is convex: From (2.47), for $\gamma_{x,k} := \gamma_x$, we have that there exists a positive number α such that:

$$\begin{aligned}
\alpha\gamma_x\mathbb{E}[\text{dist}(x^k, X^*)] &\leq \gamma_x\mathbb{E}[(x^k - x^*)^T F(x^k;\theta^*)] \\
&\leq a_k - a_{k+1} + \frac{1}{2}\gamma_x^2 M^2 - \gamma_x\mathbb{E}[(x^k - x^*)^T(F(x^k;\theta^k) - F(x^k;\theta^*))] \\
&\leq a_k - a_{k+1} + \frac{1}{2}\gamma_x^2 M^2 + \frac{1}{2}\gamma_x^{2-\tau}\mathbb{E}[\|x^k - x^*\|^2] + \frac{1}{2}\gamma_x^{\tau}\mathbb{E}[\|F(x^k;\theta^k) - F(x^k;\theta^*)\|^2] \\
&\leq a_k - a_{k+1} + \frac{1}{2}\gamma_x^2 M^2 + \frac{1}{2}\gamma_x^{2-\tau} M_x^2 + \frac{1}{2}\gamma_x^{\tau} L_\theta^2\mathbb{E}[\|\theta^k - \theta^*\|^2],
\end{aligned}$$

where $0 < \tau < 1$. It follows that

$$\begin{aligned}
\alpha\gamma_x\mathbb{E}[\text{dist}(x^k, X^*)] &\leq \limsup_{k\to\infty} a_k - \limsup_{k\to\infty} a_{k+1} + \frac{1}{2}\gamma_x^2 M^2 + \frac{1}{2}\gamma_x^{2-\tau} M_x^2 + \frac{1}{2}\gamma_x^{\tau} L_\theta^2 \limsup_{k\to\infty}\mathbb{E}[\|\theta^k - \theta^*\|^2] \\
&\leq \frac{1}{2}\gamma_x^2 M^2 + \frac{1}{2}\gamma_x^{2-\tau} M_x^2 + \frac{1}{2}\gamma_x^{\tau} L_\theta^2 \frac{\gamma_\theta \nu_\theta^2}{2\mu_\theta - \gamma_\theta C_\theta^2}.
\end{aligned}$$

It follows that

$$\limsup_{k\to\infty}\mathbb{E}[\text{dist}(x^k, X^*)] \leq \frac{1}{\alpha}\left[\frac{1}{2}\gamma_x M^2 + \frac{1}{2}\gamma_x^{1-\tau} M_x^2 + \frac{1}{2}\gamma_x^{\tau-1} L_\theta^2 \frac{\gamma_\theta \nu_\theta^2}{2\mu_\theta - \gamma_\theta C_\theta^2}\right].$$

∎

2.4 Numerical results

In this section, we apply the developed algorithms on a class of misspecified economic dispatch problems described in Section 2.4.1. In Section 2.4.2, we apply the proposed schemes for purposes of learning optimal solutions and the misspecified parameters. Note that the simulations were carried out on Tomlab 7.4. The complementarity solver `PATH` [53] was utilized for obtaining solutions to these problems which subsequently formed the basis for comparison.

2.4.1 Problem description

We consider a setting where there are N firms competing over a W-node network. Firm f may produce and sell its good at node i, where $f = 1, \ldots, N$ and $i = 1, \ldots, W$. We assume that for a given firm f, the cost of generating x_{fi} units of power at node i is random and is given by $c_{fi}(x_{fi}) = d_{fi}x_{fi}^2 + h_{fi}x_{fi} + \xi_{fi}$, where d_{fi} and h_{fi} are positive parameters, and ξ_{fi} is a random variable with mean zero for all f and i. Furthermore, the generation level associated with firm f is bounded by its production capacity, which is denoted by cap_{fi}.

The aggregate sales of all firms at node i has to satisfy the demand D_i at node i. A given firm can produce at any node and then sell at different nodes, provided that the aggregate production at all nodes matches the aggregate sales at all nodes for each firm. For simplicity, we assume that there is no limit of sales at any node. Then, the resulting problem faced by the grid operator can be stated as follows:

$$\begin{aligned} \min_{x_{fi} \geq 0} \quad & \mathbb{E}\left[\sum_{f=1}^{N}\sum_{i=1}^{W} c_{fi}(x_{fi})\right] \\ \text{subject to} \quad & x_{fi} \leq \text{cap}_{fi}, \qquad \text{for all } f, i \\ & \sum_{f=1}^{N} x_{fi} = D_i. \end{aligned} \tag{2.51}$$

The resulting optimal solution is given by x^*. Suppose firm f generates y_{fi} units of power at node i. We use $c_{fi}(y_{fi}) = d_{fi}(y_{fi})^2 + h_{fi}y_{fi} + \xi_{fi}$ to denote the cost associated with firm f at node i. The operator will solve the following (regularized) problem to estimate c_{fi} and d_{fi}:

$$\min_{\{d_{fi}, h_{f,i}\} \in \Theta} \quad \mathbb{E}\left[(d_{fi}(y_{fi})^2 + h_{fi}y_{fi} - c_{fi}(y_{fi}))^2 + \mu_\theta d_{fi}^2 + \mu_\theta h_{fi}^2\right]. \tag{2.52}$$

The resulting optimal solution is given by θ^*. We assume that y_{fi} is distributed as per a uniform distribution and is specified by $y_{fi} \sim U[0, \text{cap}_{fi}]$, while that the noise ξ_{fi} is distributed as per a uniform distribution and is specified by $\xi_{fi} \sim U[-\theta^*_{fi}/2, \theta^*_{fi}/2]$.

2.4.2 Results

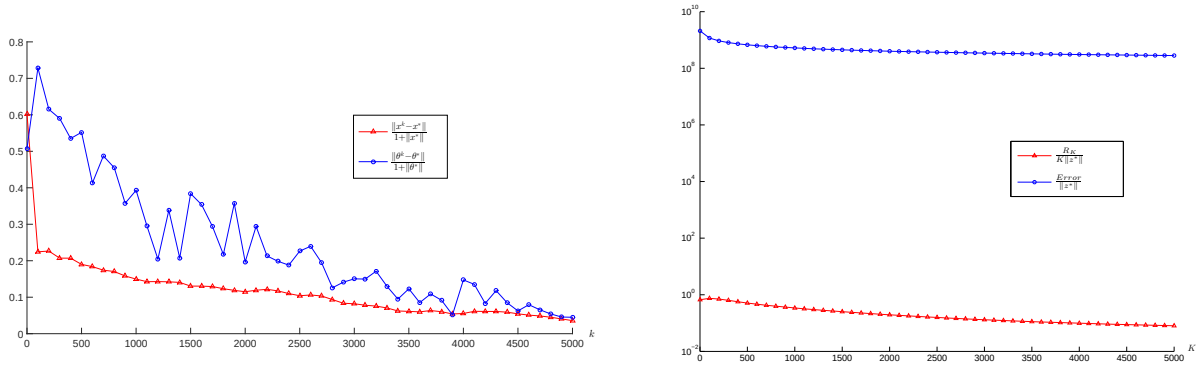
In this subsection, we employ Algorithm 1 proposed in Section 2.2 for learning parameters and computing optimal solutions. We will examine the behavior and error bounds of the algorithm.

Behavior of the algorithm

In this part, we consider a special case when $N = 5$ and $W = 5$. Suppose, the noise ξ is distributed as per a uniform distribution and is specified by $\xi \sim U[-\theta^*/2, \theta^*/2]$. Suppose the steplength sequences $\{\gamma_{k,x}\}$ and $\{\gamma_{k,\theta}\}$ are chosen according to Proportion 2: $\gamma_{k,x} = 1/k$ and $\gamma_{k,\theta} = 40/k$. Figure 2.1(a) illustrates the scaled error of the learning scheme when the number of steps increases.

Error bounds

In this part, we examine the errors of the algorithm and compare them with the theoretical error bounds proposed in Section 2.2. Suppose, the noise ξ is distributed as per a uniform distribution and is specified by



(a) Normalized error

(b) Normalized regret

Figure 2.1: Computing x^* and learning θ^* ($\xi \sim U[-\theta^*/2, \theta^*/2]$, $N = 5$, $W = 5$)

$\xi \sim U[-\theta^*/2, \theta^*/2]$.

(a) In the strongly convex regime, suppose the steplength sequences $\{\gamma_{k,x}\}$ and $\{\gamma_{k,\theta}\}$ are chosen according to Proportion 2: $\gamma_{k,x} = 1/k$ and $\gamma_{k,\theta} = 40/k$. We use ERR to denote the theoretical error provided in Proportion 2. The algorithm was terminated at $K = 10000$. Table 2.1 (L) shows the scaled errors of the learning scheme.

(b) In the merely convex regime, suppose the steplength γ_x and the steplength sequence $\{\gamma_{k,\theta}\}$ are chosen according to Theorem 2: γ_x is chosen by Table 2.1 (R) and $\gamma_{k,\theta} = 40/k$. We use ERR to denote the theoretical error provided in Theorem 2 while z^* denotes $f(x^*; \theta^*)$. The algorithm was terminated at $K = 10000$ and Table 2.1(R) shows the scaled errors of the learning scheme.

(c) Suppose the steplength sequences $\{\gamma_{k,x}\}$ and $\{\gamma_{k,\theta}\}$ are chosen according to Theorem 4: $\gamma_{k,x} = k^{-\alpha}$ and $\gamma_{k,\theta} = 40/k$. We employ ERR to denote the theoretical error provided in Theorem 4 while z^* denotes $f(x^*; \theta^*)$. The algorithm was terminated after $K = 10000$ iterations. Figure 2.1(b) illustrates the scaled regret and scaled theoretical error of the learning scheme when the number of steps increases ($\alpha = \beta = 0.5$). Table 2.2 shows the scaled theoretical error of the learning scheme for different chosen $\gamma_{k,x} = k^{-\alpha}$ with $\alpha = 0.5, 0.6, 0.7, 0.8, 0.9$ when $\beta = 0.5$. We see that when α changes, error bounds change marginally primarily because the last term in Theorem 4 dominates the bound.

Table 2.1: Learning x^* and θ^* in a strongly convex (L) and convex (R) regime: $\xi \sim U[-\theta^*/2, \theta^*/2]$

N	W	$\frac{\mathbb{E}[\|x^K - x^*\|]}{1+\|x^*\|}$	$\frac{\text{ERR}}{1+\|x^*\|}$	$\frac{\|\mathbb{E}[\theta^K - \theta^*\|]}{1+\|\theta^*\|}$	$\frac{\text{ERR}}{1+\|\theta^*\|}$
10	2	7.3×10^{-3}	9.2×10^{9}	4.8×10^{-2}	3.7×10^{4}
10	4	3.7×10^{-2}	2.1×10^{10}	4.9×10^{-2}	3.1×10^{4}
10	6	3.8×10^{-2}	7.8×10^{10}	4.7×10^{-2}	8.3×10^{4}
10	8	1.7×10^{-2}	9.1×10^{10}	4.8×10^{-2}	8.5×10^{4}
10	10	2.4×10^{-2}	1.2×10^{11}	4.3×10^{-2}	8.6×10^{4}

N	W	$\frac{\mathbb{E}[f(\tilde{x}_{1,K}; \theta^K) - z^*]}{1+\|z^*\|}$	$\frac{\text{ERR}}{1+\|x^*\|}$	γ_x
10	2	1.9×10^{-1}	2.5×10^{5}	72
10	4	6.5×10^{-2}	1.1×10^{5}	93
10	6	2.7×10^{-1}	2.6×10^{5}	127
10	8	1.3×10^{-1}	1.7×10^{5}	131
10	10	1.4×10^{-1}	2.6×10^{5}	133

Table 2.2: Investigation of regret when learning x^* and θ^* in a stochastic convex regime: $\xi \sim U[-\theta^*/2, \theta^*/2]$, $N = 5$, $W = 5$

α	$\frac{R_K}{K\|z^*\|}$	$\frac{\text{ERR}}{\|z^*\|}$
0.5	4.8×10^{-2}	3.1×10^{8}
0.6	3.3×10^{-2}	3.1×10^{8}
0.7	2.3×10^{-2}	3.1×10^{8}
0.8	1.8×10^{-2}	3.1×10^{8}
0.9	1.5×10^{-2}	3.1×10^{3}

2.5 Concluding remarks

Traditionally, much of the field of optimization has been defined by problems in which the functions and sets are known to the decision-maker. However, as problems grow in their reliance on data, such knowledge cannot be taken for granted. We consider one such instance of such problems where functions may be misspecified and the associated vector may be learnt through the parallel solution of a suitably defined problem. It is worth emphasizing the problem in the *full* space of learning and optimization variables is a challenging (non-monotone) stochastic variational problem for which no first-order methods are currently available. Yet, by leveraging the structure of the problem, we show that such problems can indeed be efficiently solved.

We consider a problem of solving a stochastic optimization problem in which the objective is parameterized by a vector that can be learnt by solving a suitably defined learning problem, captured by a stochastic optimization problem. In both strongly convex and merely convex regimes, we develop a set of coupled stochastic approximation schemes which produces a sequence of iterates that are shown to converge to the solution and unknown parameter in an almost sure sense. Additionally, we provide rate estimates for the prescribed schemes in both strongly convex and convex regimes. Through an analysis of the rate of convergence under a diminishing steplength setting, it is seen that the optimal rate of convergence is observed in strongly convex problems while in convex regimes, we see a degradation introduced by learning from $\mathcal{O}\left(\frac{1}{\sqrt{K}}\right)$ to $\mathcal{O}\left(\frac{\sqrt{\ln(K)}}{\sqrt{K}}\right)$. This degradation is seen to disappear if the averaging window is modified appropriately. Similar rate statements are also provided in a constant steplength regime. In fact, we may also cast this problem as an online decision-making problem where a decision-maker sees a collection of misspecified functions. In a stochastic regime, we observe that an upper bound on the average regret can be shown to decay at a rate no worse than $\mathcal{O}\left(\frac{\ln K}{\sqrt{K}}\right)$ for a suitably chosen steplength.

Unfortunately, the optimization-based model cannot accommodate settings where there is misspecification in the constraints or, more generally, if the associated decision-making problem is an equilibrium problem. Motivated by this gap, we consider a misspecified stochastic variational inequality problem and propose analogous stochastic approximation schemes for computation and learning. To resolve the challenge associated with merely monotone maps, we employ (Tikhonov) regularized counterparts for which almost-sure convergence statements can be provided. Additionally, we provide rate statements for constant and

diminishing steplength regimes, of which the latter requires imposing a suitable weak-sharpness assumption on the original problem. Again, it is seen that while the schemes display the optimal rate of convergence under strongly monotone regimes, a degradation in the rate is seen in the monotone regime.

Chapter 3

Misspecified Stochastic Nash Games

3.1 Introduction

In networked engineered systems, a common challenge lies in designing distributed control architectures that ensure the satisfaction of a system-wide criterion in environments complicated by nonlinearity, uncertainty, and dynamics. Such control-theoretic problems take on a variety of forms and arise in a variety of networked settings, including networks of unmanned aerial vehicles (UAVs), traffic networks, wireline and wireless communication networks, and energy systems, amongst others. These systems are often characterized by the absence of a designated central entity that either has system-wide control or has access to global information. Consequently, control is effected through distributed decision-making and local interactions that rely on limited information. Game-theoretic approaches represent an avenue for designing such protocols. Game theory has seen wide applicability in the social, economic, and engineered sciences in a largely *descriptive* role. There has been immense recent interest in a *prescriptive role* [54] that considers *designing a game* whose equilibria represent the solutions to the control problem of interest [55, 56]; consequently, the distributed learning of Nash equilibria assumes immediate relevance in the management of networked systems. Learning in Nash games has seen much study in the last several decades [57, 58, 59, 60]. In continuous strategy regimes, convex static games find significance in engineered systems such as communication networks [61, 62, 63, 64] and signal processing [65, 66].

An oft-used assumption in game-theoretic models requires that player payoffs are public knowledge, allowing every player to correctly forecast the choices of his adversaries. As noted by Kirman [67], a firm's view of the game may be corrupted or *misspecified* in at least two distinct ways in a Cournot setting where firms decide production levels given a price function: (i) a firm might have a correct description of the price function but an incorrect estimate of its parameters; and (ii) it may have an incorrect structure of the price function and incorrectly conclude that prediction errors are a consequence of misspecified parameters. Kirman [67] considered such a learning process, and showed that by observing true demand, the suggested learning process guarantees that the firm strategies converge to the noncooperative Nash equilibrium. Further

inspiration may be drawn from studies by Bischi [68, 69], Szidarovsky [70, 71], and others [72], where firms competing in a deterministic Nash-Cournot game learn a parameter of the demand function while playing the game repeatedly. Note that an inherent assumption of a low discount rate is imposed that discounts the future effect of any player's strategies. Analogous questions of optimization and estimation have also been studied by Cooper et al. [38] who consider a joint process of forecasting and optimization in a regime where the underlying model may be erroneous, demonstrating that the resulting revenues can systematically reduce over time.

When designing protocols for practical engineered systems, particularly in the absence of a centralized controller, the associated parameters of the utility functions may often be misspecified. For instance, in power market models that enlist Nash-Cournot models [73, 74], the precise nature of the price function is assumed to be given. Similarly, the expected capacity or availability of renewable generation assets is rarely known a priori. Similarly, when developing distributed protocols for networked UAVs, the prescribed utility functions may rely on agent-specific information that can only be learnt through observations. Faced by such challenges, our goal lies in the development and analysis of general purpose algorithms that combine computation of Nash equilibria with a learning phase to correct misspecification.

Motivation: This chapter is motivated by the absence of general-purpose distributed schemes with asymptotic convergence and rate guarantees for learning equilibria in the face of imperfect information. Such problems emerge from stochastic generalizations of problems arising in communication networks [62, 75, 63, 64], signal processing [65, 66], and power markets [73]. Accordingly, we present two distributed learning schemes in which agents *learn their Nash strategy* while *correcting the misspecification* in their payoffs:

(1) **Stochastic gradient schemes for stochastic Nash games:** In Section 3.2, we present a distributed stochastic approximation framework in which every agent makes two projected gradient updates: Every agent first updates its belief regarding the equilibrium strategy by using the sampled gradient of its payoff function and subsequently updates its belief regarding the misspecified parameter through a similar projected (stochastic) gradient update. The resulting sequence of equilibrium and parameter estimates are shown to converge to their true counterparts in an almost sure sense. Notably, we show that the mean-squared error of the equilibrium estimates converges to zero at the optimal rate $\mathcal{O}(1/K)$ despite the presence of misspecification.

(2) **Iterative fixed-point schemes for stochastic Nash-Cournot games:** In Section 3.3, we consider a Cournot regime where aggregate output is unobservable and one parameter of the demand function is misspecified. Under common-knowledge, agents develop an estimate of aggregate output and the misspecified price function parameter by observing noisy prices. These estimates allow developing an iterative fixed-point scheme that produces iterates that are shown to converge to the Nash-Cournot equilibrium in an almost-sure

sense. Additionally, firms learn the true parameter in an almost-sure sense. The result can be extended to nonlinear price functions.

Remark: We make two remarks before proceeding. (a) First, in (1), the learning problem is constructed independently of the computational problem through a set of observations while in (2), the learning is affected by the computational step (akin to multi-armed bandit problems). (b) Second, we comment on the sequential two-stage framework for resolving misspecification:

$$\text{Step 1. Learn } \theta^* \qquad \text{Step 2. Compute } x^*(\theta^*),$$

where θ^* is to be learnt and $x^*(\theta^*)$ is the (stochastic) Nash equilibrium, given θ^*. Unfortunately, such an approach is complicated by several challenges. First, Step 1. needs to be completed in a finite number of iterations, practically impossible for stochastic learning problems. Second, premature termination of Step 1. leads to an erroneous estimate $\hat{\theta}$ leading to an incorrect Nash equilibrium $\hat{x}$. In fact, in stochastic regimes, one often cannot prescribe the amount of learning effort required in a priori sense. Preliminary numerics reveal that sequential schemes may perform orders of magnitude worse when compared with iterative fixed-point schemes (see Table 3.3). Third, offline or a priori observations may be unavailable as required by Step 1.

This chapter is organized as follows. In Section 3.2, we define and resolve a misspecified stochastic Nash game and present a joint set of stochastic approximation schemes that jointly allow for learning equilibria and resolving misspecification. In Section 3.3, we develop iterative fixed-point schemes in Cournot settings where aggregate output is unobservable. Empirical studies and conclusions are provided in Sections 3.4 and 3.5, respectively.

3.2 Gradient-based schemes for convex Nash games

3.2.1 Problem description, assumptions and background

We consider an $N-$person stochastic Nash game in which the ith player solves $\text{Opt}(x_{-i})$:

$$\min_{x_i \in K_i} \quad f_i(x;\theta^*) \triangleq \mathbb{E}[f_i(x;\theta^*,\xi)] \tag{$\text{Opt}(x_{-i})$}$$

where $K_i \subseteq \mathbb{R}^{n_i}$, $\theta^* \in \mathbb{R}^m$, $\xi : \Omega \to \mathbb{R}^d$ defined on a probability space $(\Omega_x, \mathcal{F}_x, \mathbb{P}_x)$, $n = \sum_{i=1}^N n_i$, and $f_i : \mathbb{R}^n \times \mathbb{R}^m \times \mathbb{R}^d \to \mathbb{R}$ is a real-valued function in x_i, $x_{-i} \triangleq (x_j)_{i\neq j=1}^N$, and ξ. The associated Nash equilibrium is given by a tuple $x^* = (x_i^*)_{i=1}^N$ where $x_i^* \in \text{SOL}(\text{Opt}(x_{-i}^*))$ for $i = 1, \ldots, N$, $\text{SOL}(\text{Opt}(x_{-i}^*))$

denotes the solution of $\text{Opt}(x_{-i})$ and under suitable convexity requirements (see (A10) below), x^* is a solution to a stochastic variational inequality problem $\text{VI}(K, F(\bullet;\theta^*))$ where K and $F : K \times \Theta \to \mathbb{R}^n$ are defined as follows:

$$K \triangleq \prod_{i=1}^{N} K_i \text{ and } F(x;\theta) \triangleq \left(\mathbb{E}[\nabla_{x_i} f_i(x;\theta,\xi)]\right)_{i=1}^{N}, \tag{3.1}$$

respectively. It may be recalled that $\text{VI}(K, F)$ requires an $x \in K$ satisfying

$$(y-x)^T F(x;\theta^*) \geq 0, \quad \text{for all } y \in K. \tag{3.2}$$

Our overall goal lies in computing equilibria when θ^* is unavailable or misspecified.

Learning scheme In this section, we consider the estimation of θ^* through the solution of a suitably defined stochastic convex learning problem [1]:

$$\min_{\theta\in\Theta} \quad g(\theta) \triangleq \mathbb{E}[g(\theta;\eta)], \tag{3.3}$$

where $\Theta \subseteq \mathbb{R}^m$ is a closed and convex set, $\eta : \mathcal{Z} \to \mathbb{R}^p$ is a random variable defined on a probability space $(\Lambda, \mathcal{F}_\theta, \mathbb{P}_\theta)$, and $g : \Theta \times \Lambda \to \mathbb{R}$ is a real-valued learning metric function (such as a regression metric constructed from a set of observations). Consequently, θ^* may be learnt through a stochastic gradient scheme of the form:

$$\theta_i^{k+1} := \Pi_\Theta\left(\theta_i^k - \alpha_i^k \nabla_\theta g(\theta_i^k;\eta_i^k)\right), \quad k \geq 0, \quad i = 1,\ldots,N. \tag{3.4}$$

We emphasize that this learning problem is unrelated to the computational process and is a built from a set of independently collected observations.

Distributed computational scheme We consider a distributed stochastic approximation scheme where the ith agent employs its belief regarding θ^* to take a (stochastic) gradient step:

$$x_i^{k+1} := \Pi_{K_i}\left(x_i^k - \gamma_i^k \nabla_{x_i} f_i(x^k;\theta^k,\xi^k)\right), \quad k \geq 0, \quad i = 1,\ldots,N, \tag{3.5}$$

where γ_i^k and $\nabla_{x_i} f_i(x^k;\theta^k,\xi^k)$ denotes the steplength and sampled gradient used by player i at step k and $\Pi_X(u)$ denotes the Euclidean projection of u onto X. While a fully rational agent would always take a best response step, in stochastic settings, the complexity of this step might be significant. In bounded

rational regimes where computational constraints are imposed, an alternative lies in computing other steps (such as the gradient-response) (cf. [76, 77]) (cf. research in communication networks [61] and cognitive radio games [78]. An alternate motivation arises from distributed control/optimization settings where a "game" is designed whose equilibrium is a desirable solution to a suitably defined control problem. Here, a distributed protocol for computing an equilibrium can be designed and gradient-based approaches can be adopted (cf. [54, 55, 56]). We propose a game-theoretic extension of that developed in [79]. We may specify our joint simulation-based scheme for learning and computation as follows:

Algorithm 4 (Gradient response and learning). Step 0. *Given* $\theta_i^0 \in \Theta$, $x_0 \in K$, $\{\gamma_i^k, \alpha_i^k\} > 0$, *for* $i = 1, \ldots, N$, *and* $k = 0$.

Step 1:

$$x_i^{k+1} := \Pi_{K_i}\left(x_i^k - \gamma_i^k \nabla_{x_i} f_i(x^k; \theta_i^k, \xi_i^k)\right), \qquad k \geq 0, \quad i = 1, \ldots, N \qquad \text{(Computation)}$$

$$\theta_i^{k+1} := \Pi_{\Theta}\left(\theta_i^k - \alpha_i^k \nabla_\theta g(\theta_i^k; \eta_i^k)\right), \qquad k \geq 0, \quad i = 1, \ldots, N. \qquad \text{(Learning)}$$

Step 2: *if* $k > \bar{K}$, *stop; else* $k := k + 1$ *and go to Step 1.*

We now present the main assumptions employed in deriving convergence properties of Algorithm 4. (A1) enforces convexity assumptions that allow for deriving sufficient equilibrium conditions as $\mathrm{VI}(X, F)$ while the monotonicity requirements on F allow for claiming the existence of a unique equilibrium. Lipschitzian requirements of F aid in deriving subsequent convergence and rate statements. Furthermore, a breadth of learning problems (such as regression, classification etc. [1]) are convex. The requirements imposed by (A11) are standard in developing distributed protocols while (A12) imposes assumptions on the conditional first and second moments common in stochastic approximation literature [80, 43, 81].

Assumption 10 (A10)**.** *For* $i = 1, \ldots, N$, *suppose the function* $f_i(x; \theta)$ *is convex and continuously differentiable function in* x_i *for every* $x_{-i} \in \prod_{j \neq i} K_j$ *and every* $\theta \in \Theta$. *Furthermore, suppose* Θ *is a closed, convex, and bounded set and for* $i = 1, \ldots, N$, $K_i \subseteq \mathbb{R}^{n_i}$ *is a nonempty, closed, convex and bounded set. Furthermore, suppose the following hold:(a) For every* $\theta \in \Theta$, $F(x; \theta)$ *is both strongly monotone and Lipschitz continuous in* x *with constants* μ_x *and* L_x*; for every* θ, $(F(x; \theta) - F(y; \theta))^T (x - y) \geq \mu_x \|x - y\|^2$, *and* $\|F(x; \theta) - F(y; \theta)\| \leq L_x \|x - y\|$*; (b) For every* $x \in K$, $F(x; \theta)$ *is Lipschitz continuous in* θ *with constant* L_θ*; (c) The function* $g(\theta)$ *is strongly convex and continuously differentiable with Lipschitz continuous gradients in* θ *with convexity constant* μ_θ *and Lipschitz constant* C_θ, *respectively.*

Note that monotone Nash games include *stable* Nash games, a class of games for which it has been shown that a range of evolutionary dynamics allow for convergence to Nash equilibria [82]. In fact, in recent

work [83], a notion of passivity has been developed.

Assumption 11 (A11). *For $i = 1, \ldots, N$, the ith agent knows only his objective f_i and strategy set K_i. Furthermore, the vector x is assumed to be observable.*

We define a new probability space $(Z, \mathcal{F}, \mathbb{P})$, where $Z \triangleq \Omega \times \Lambda$, $\mathcal{F} \triangleq \mathcal{F}_x \times \mathcal{F}_\theta$ and $\mathbb{P} \triangleq \mathbb{P}_x \times \mathbb{P}_\theta$. For $i = 1, \ldots, N$, suppose $w_i^k \triangleq \nabla_{x_i} f_i(x^k; \theta_i^k, \xi^k) - \nabla_{x_i} f_i(x^k; \theta_i^k)$ and $v_i^k \triangleq \nabla_\theta g(\theta_i^k; \eta^k) - \nabla_\theta g(\theta_i^k)$. $\mathcal{F}_k$ denotes the sigma-field generated by (x^0, θ^0) and errors (w^l, v^l) for $l = 0, 1, \cdots, k-1$, i.e., $\mathcal{F}_0 = \sigma\left\{(x^0, \theta^0)\right\}$ and $\mathcal{F}_k = \sigma\left\{(x^0, \theta^0), \left((w^l, v^l), l = 0, 1, \cdots, k-1\right)\right\}$ for $k \geq 1$.

Assumption 12 (A12). *(a) Unbiasedness: $\mathbb{E}[w^k \mid \mathcal{F}_k] = 0$ and $\mathbb{E}[v_i^k \mid \mathcal{F}_k] = 0$ a.s. for all k and i; (b)Bounded second moments: $\mathbb{E}[\|w^k\|^2 \mid \mathcal{F}_k] \leq \nu_x^2$ and $\mathbb{E}[\|v_i^k\|^2 \mid \mathcal{F}_k] \leq \nu_\theta^2$ a.s. for all k, i.*

To construct distributed schemes requiring no coordination in terms of setting parameters, we allow each agent to independently set steplengths and as long as a suitable relationship between these steplengths holds, convergence follows. Specifically, the ith agent employs a diminishing steplength sequence given by γ_i^k. Furthermore, we define $\gamma_{\min}^k \triangleq \min_{1 \leq i \leq N}\{\gamma_i^k\}$ and $\gamma_{\max}^k \triangleq \max_{1 \leq i \leq N}\{\gamma_i^k\}$ for all k. Similarly, we define $\alpha_{\min}^k \triangleq \min_{1 \leq i \leq N}\{\alpha_i^k\}$ and $\alpha_{\max}^k \triangleq \max_{1 \leq i \leq N}\{\alpha_i^k\}$ for all k. Then, we can make the following assumptions on the steplengths of the algorithm.

Assumption 13 (Steplength requirements, A13). *Let $\{\gamma_i^k\}$ and $\{\alpha_i^k\}$ be chosen such that: (a) $\sum_{k=1}^\infty \gamma_{\min}^k = \infty$, $\sum_{k=1}^\infty (\gamma_{\max}^k)^2 < \infty$, $\sum_{k=1}^\infty (\alpha_{\max}^k)^2 < \infty$; (b) $\lim_{k\to\infty} \frac{\gamma_{\max}^k - \gamma_{\min}^k}{\gamma_{\max}^k} = 0$; (c) $\alpha_{\min}^k \geq \gamma_{\max}^k L_\theta^2/(\mu_x \mu_\theta)$ for sufficiently large k, $\lim_{k\to\infty} \frac{(\alpha_{\max}^k)^2}{\gamma_{\max}^k} = 0$.*

Notice that (a)$\sum_{k=1}^\infty \gamma_{\min}^k = \infty$ and (c) $\alpha_{\min}^k \geq \gamma_{\max}^k L_\theta^2/(\mu_x \mu_\theta)$ for sufficiently large k implies that $\sum_{k=1}^\infty \alpha_{\min}^k = \infty$.

3.2.2 Analysis

We begin with a contraction statement for the sequence of iterates produced by Algorithm 4.

Lemma 8. *Suppose (A10), (A11), (A12) and (A13) hold. Let $\{x^k, \theta^k\}$ be computed via Algorithm 4. For any $k \geq 0$, $\mathbb{E}\left[\|x^{k+1} - x^*\|^2 \mid \mathcal{F}_k\right] \leq \zeta_k \|x^k - x^*\|^2 + \beta_k$, where $\zeta_k = 1 - \gamma_{\max}^k \mu_x + 2(\gamma_{\max}^k - \gamma_{\min}^k) L_x + 2(\gamma_{\max}^k)^2 L_x^2$ and $\beta_k = (2(\gamma_{\max}^k)^2 L_\theta^2 + \gamma_{\max}^k L_\theta^2/\mu_x) \sum_{i=1}^N \|\theta_i^k - \theta^*\|^2 + (\gamma_{\max}^k)^2 \nu_x^2$.*

Proof. Since $x_i^* = \Pi_{K_i}(x_i^* - \gamma_i^k F_i(x^*;\theta^*))$, by the nonexpansivity of the Euclidean projector:

$$\begin{aligned}\|x^{k+1} - x^*\|^2 \leq \sum_{i=1}^N (\|x_i^k - x_i^*\|^2 + (\gamma_i^k)^2\|F_i(x^k;\theta_i^k) - F_i(x^*;\theta^*)\|^2 + (\gamma_i^k)^2\|w_i^k\|^2) \\ - 2\sum_{i=1}^N \gamma_i^k (x_i^k - x_i^*)^T (F_i(x^k;\theta_i^k) - F_i(x^*;\theta^*)) \\ - 2\sum_{i=1}^N \gamma_i^k (x_i^k - x_i^*)^T w_i^k + 2\sum_{i=1}^N (\gamma_i^k)^2 (F_i(x^k;\theta_i^k) - F_i(x^*;\theta^*))^T w_i^k. \end{aligned} \tag{3.6}$$

$$\begin{aligned}\text{RHS of (3.6)} \leq \underbrace{\|x^k - x^*\|^2 + (\gamma_{\max}^k)^2 \sum_{i=1}^N \|F_i(x^k;\theta_i^k) - F_i(x^*;\theta^*)\|^2}_{\textbf{term 1}} + (\gamma_{\max}^k)^2\|w^k\|^2 \\ \underbrace{-2\sum_{i=1}^N \gamma_i^k (x_i^k - x_i^*)^T (F_i(x^k;\theta_i^k) - F_i(x^k;\theta^*))}_{\textbf{term 2}} \underbrace{-2\sum_{i=1}^N \gamma_i^k (x_i^k - x_i^*)^T (F_i(x^k;\theta^*) - F_i(x^*;\theta^*))}_{\textbf{term 3}} \\ - 2\sum_{i=1}^N \gamma_i^k (x_i^k - x_i^*)^T w_i^k + 2\sum_{i=1}^N (\gamma_i^k)^2 (F_i(x^k;\theta_i^k) - F_i(x^*;\theta^*))^T w_i^k. \end{aligned} \tag{3.7}$$

By (A10), term 1 in (3.7) may be bounded by leveraging the Lipschitz continuity of $F(x;\theta)$:

$$\begin{aligned}&\|x^k - x^*\|^2 + 2(\gamma_{\max}^k)^2 \sum_{i=1}^N \|F_i(x^k;\theta_i^k) - F_i(x^k;\theta^*)\|^2 + 2(\gamma_{\max}^k)^2 \sum_{i=1}^N \|F_i(x^k;\theta^*) - F_i(x^*;\theta^*)\|^2 \\ &\leq \|x^k - x^*\|^2 + 2(\gamma_{\max}^k)^2 \sum_{i=1}^N \|F(x^k;\theta_i^k) - F(x^k;\theta^*)\|^2 + 2(\gamma_{\max}^k)^2 \|F(x^k;\theta^*) - F(x^*;\theta^*)\|^2 \\ &\leq (1 + 2(\gamma_{\max}^k)^2 L_x^2)\|x^k - x^*\|^2 + 2(\gamma_{\max}^k)^2 L_\theta^2 \sum_{i=1}^N \|\theta_i^k - \theta^*\|^2. \end{aligned} \tag{3.8}$$

By (A10), term 2 in (3.7) can be bounded by the Cauchy-Schwarz inequality, Hölder's inequality and the Lipschitz continuity of $F(x;\theta)$:

$$\begin{aligned}&-2\sum_{i=1}^N \gamma_i^k (x_i^k - x_i^*)^T (F_i(x^k;\theta_i^k) - F_i(x^k;\theta^*)) \leq 2\gamma_{\max}^k \sum_{i=1}^N \|x_i^k - x_i^*\| \|F_i(x^k;\theta_i^k) - F_i(x^k;\theta^*)\| \\ &\leq 2\gamma_{\max}^k \|x^k - x^*\| \sqrt{\sum_{i=1}^N \|F_i(x^k;\theta_i^k) - F_i(x^k;\theta^*)\|^2} \leq 2\gamma_{\max}^k \|x^k - x^*\| \sqrt{\sum_{i=1}^N \|F(x^k;\theta_i^k) - F(x^k;\theta^*)\|^2} \\ &\leq 2\gamma_{\max}^k L_\theta \|x^k - x^*\| \sqrt{\sum_{i=1}^N \|\theta_i^k - \theta^*\|^2} \leq \gamma_{\max}^k \mu_x \|x^k - x^*\|^2 + \gamma_{\max}^k \frac{L_\theta^2}{\mu_x} \sum_{i=1}^N \|\theta_i^k - \theta^*\|^2, \end{aligned} \tag{3.9}$$

where the last inequality follows from the fact that $2ab \leq a^2 + b^2$. Term 3 in (3.7) can be bounded by the

Cauchy-Schwarz inequality and $\gamma_i^k \leq \gamma_{\max}^k$ for all i:

$$
\begin{aligned}
&-2\sum_{i=1}^{N}\gamma_i^k(x_i^k-x_i^*)^T(F_i(x^k;\theta^*)-F_i(x^*;\theta^*)) = -2\sum_{i=1}^{N}\gamma_{\max}^k(x_i^k-x_i^*)^T(F_i(x^k;\theta^*)-F_i(x^*;\theta^*))\\
&-2\sum_{i=1}^{N}(\gamma_i^k-\gamma_{\max}^k)(x_i^k-x_i^*)^T(F_i(x^k;\theta^*)-F_i(x^*;\theta^*))\\
&\leq -2\gamma_{\max}^k\sum_{i=1}^{N}(x_i^k-x_i^*)^T(F_i(x^k;\theta^*)-F_i(x^*;\theta^*))+2(\gamma_{\max}^k-\gamma_{\min}^k)\sum_{i=1}^{N}\|x_i^k-x_i^*\|\|F_i(x^k;\theta^*)-F_i(x^*;\theta^*)\|.
\end{aligned}
$$

Proceeding further, we may leverage Hölder's inequality, the Lipschitz continuity of $F(x;\theta)$ and (A10), to obtain the following sequence of inequalities:

$$
\begin{aligned}
&-2\gamma_{\max}^k\sum_{i=1}^{N}(x_i^k-x_i^*)^T(F_i(x^k;\theta^*)-F_i(x^*;\theta^*))+2(\gamma_{\max}^k-\gamma_{\min}^k)\sum_{i=1}^{N}\|x_i^k-x_i^*\|\|F_i(x^k;\theta^*)-F_i(x^*;\theta^*)\|\\
&\leq -2\gamma_{\max}^k(x^k-x^*)^T(F(x^k;\theta^*)-F(x^*;\theta^*))+2(\gamma_{\max}^k-\gamma_{\min}^k)\|x^k-x^*\|\|F(x^k;\theta^*)-F(x^*;\theta^*)\|\\
&\leq -2\gamma_{\max}^k\mu_x\|x^k-x^*\|^2+2(\gamma_{\max}^k-\gamma_{\min}^k)L_x\|x^k-x^*\|^2. \qquad (3.10)
\end{aligned}
$$

Combining (3.6) with (3.7), (3.8), (3.9), and (3.10), we obtain

$$
\begin{aligned}
\|x^{k+1}-x^*\|^2 &\leq (1+2(\gamma_{\max}^k)^2L_x^2)\|x^k-x^*\|^2-\gamma_{\max}^k\mu_x\|x^k-x^*\|^2+2(\gamma_{\max}^k-\gamma_{\min}^k)L_x\|x^k-x^*\|^2\\
&\quad+2(\gamma_{\max}^k)^2L_\theta^2\sum_{i=1}^{N}\|\theta_i^k-\theta^*\|^2+\gamma_{\max}^kL_\theta^2/\mu_x\sum_{i=1}^{N}\|\theta_i^k-\theta^*\|^2+(\gamma_{\max}^k)^2\|w^k\|^2\\
&\quad-2\sum_{i=1}^{N}\gamma_i^k(x_i^k-x_i^*)^Tw_i^k+2\sum_{i=1}^{N}(\gamma_i^k)^2(F_i(x^k;\theta_i^k)-F_i(x^*;\theta^*))^Tw_i^k\\
&=\left(1-\gamma_{\max}^k\mu_x+2(\gamma_{\max}^k-\gamma_{\min}^k)L_x+2(\gamma_{\max}^k)^2L_x^2\right)\|x^k-x^*\|^2\\
&\quad+(2(\gamma_{\max}^k)^2L_\theta^2+\gamma_{\max}^kL_\theta^2/\mu_x)\sum_{i=1}^{N}\|\theta_i^k-\theta^*\|^2+(\gamma_{\max}^k)^2\|w^k\|^2\\
&\quad-2\sum_{i=1}^{N}\gamma_i^k(x_i^k-x_i^*)^Tw_i^k+2\sum_{i=1}^{N}(\gamma_i^k)^2(F_i(x^k;\theta_i^k)-F_i(x^*;\theta^*))^Tw_i^k.
\end{aligned}
$$

By taking conditional expectations and by recalling that $\mathbb{E}[w^k \mid \mathcal{F}_k]=0$ and $\mathbb{E}[\|w^k\|^2 \mid \mathcal{F}_k]\leq \nu_x^2$, we obtain that $\mathbb{E}\left[\|x^{k+1}-x^*\|^2 \mid \mathcal{F}_k\right]\leq \zeta_k\|x^k-x^*\|^2+\beta_k$, where $\zeta_k = 1-\gamma_{\max}^k\mu_x+2(\gamma_{\max}^k-\gamma_{\min}^k)L_x+2(\gamma_{\max}^k)^2L_x^2$ and $\beta_k = (2(\gamma_{\max}^k)^2L_\theta^2+\gamma_{\max}^kL_\theta^2/\mu_x)\sum_{i=1}^{N}\|\theta_i^k-\theta^*\|^2+(\gamma_{\max}^k)^2\nu_x^2$. ∎

We may now prove our main a.s. convergence result for the sequences $\{x^k\}$ and $\{\theta^k\}$.

Theorem 8. *Suppose (A10), (A11), (A12) and (A13) hold. Let $\{x^k,\theta^k\}$ be computed via Algorithm 4. Then, $x^k \xrightarrow{a.s.} x^*$ and $\theta_i^k \xrightarrow{a.s.} \theta^*$ as $k\to\infty$ for all i.*

Proof. From Lemma 8, the following holds for every k:

$$\begin{aligned}\mathbb{E}\left[\|x^{k+1}-x^*\|^2 \mid \mathcal{F}_k\right] \leq & \underbrace{\left(1-\gamma_{\max}^k\mu_x+2(\gamma_{\max}^k-\gamma_{\min}^k)L_x+2(\gamma_{\max}^k)^2L_x^2\right)}_{\triangleq\,\zeta_k}\|x^k-x^*\|^2 \\ & \underbrace{+\left(2(\gamma_{\max}^k)^2L_\theta^2+\gamma_{\max}^kL_\theta^2/\mu_x\right)\sum_{i=1}^N\|\theta_i^k-\theta^*\|^2+(\gamma_{\max}^k)^2\nu_x^2\,.}_{\triangleq\,\beta_k}\end{aligned} \tag{3.11}$$

By invoking the fixed-point property given by $\theta^* = \Pi_\Theta(\theta^* - \alpha_i^k\nabla_\theta g(\theta^*))$ (see [20]) and the non-expansivity of the Euclidean projector, we may derive the following bound on $\|\theta_i^{k+1}-\theta^*\|^2$:

$$\begin{aligned}&\|\theta_i^{k+1}-\theta^*\|^2 \leq \|\theta_i^k-\theta_i^*-\alpha_i^k(\nabla_\theta g(\theta_i^k)-\nabla_\theta g(\theta^*))-\alpha_i^kv_i^k\|^2 \\ &= \|\theta_i^k-\theta_i^*-\alpha_i^k(\nabla_\theta g(\theta_i^k)-\nabla_\theta g(\theta^*))\|^2+(\alpha_i^k)^2\|v_i^k\|^2-2\alpha_i^k(\theta_i^k-\theta_i^*-\alpha_i^k(\nabla_\theta g(\theta_i^k)-\nabla_\theta g(\theta^*)))^Tv_i^k.\end{aligned}$$

By taking conditional expectations, recalling that $\mathbb{E}[v_i^k \mid \mathcal{F}_k] = 0$ and using Lemma 4, we obtain the following bound:

$$\begin{aligned}\mathbb{E}[\|\theta_i^{k+1}-\theta^*\|^2 \mid \mathcal{F}_k] &\leq \|\theta_i^k-\theta^*-\alpha_i^k(\nabla_\theta g(\theta_i^k)-\nabla_\theta g(\theta^*))\|^2+(\alpha_i^k)^2\mathbb{E}[\|v_i^k\|^2 \mid \mathcal{F}_k] \\ &\leq (1-2\alpha_i^k\mu_\theta+(\alpha_i^k)^2C_\theta^2)\|\theta_i^k-\theta^*\|^2+(\alpha_i^k)^2\nu_\theta^2 \\ &\leq (1-2\alpha_{\min}^k\mu_\theta+(\alpha_{\max}^k)^2C_\theta^2)\|\theta_i^k-\theta^*\|^2+(\alpha_{\max}^k)^2\nu_\theta^2.\end{aligned} \tag{3.12}$$

Next, by adding (3.11) and (3.12) and by invoking (A13), we obtain the following bound.

$$\begin{aligned}&\mathbb{E}\left[\|x^{k+1}-x^*\|^2 \mid \mathcal{F}_k\right]+\mathbb{E}\left[\sum_{i=1}^N\|\theta_i^{k+1}-\theta^*\|^2 \mid \mathcal{F}_k\right] \\ &\leq \left(1-\gamma_{\max}^k\mu_x+2(\gamma_{\max}^k-\gamma_{\min}^k)L_x+2(\gamma_{\max}^k)^2L_x^2\right)\|x^k-x^*\|^2 \\ &+(1-2\alpha_{\min}^k\mu_\theta+(\alpha_{\max}^k)^2C_\theta^2+2(\gamma_{\max}^k)^2L_\theta^2+\gamma_{\max}^kL_\theta^2/\mu_x)\sum_{i=1}^N\|\theta_i^k-\theta^*\|^2+(\gamma_{\max}^k)^2\nu_x^2+N(\alpha_{\max}^k)^2\nu_\theta^2 \\ &\leq (1-\gamma_{\max}^k\mu_x+2(\gamma_{\max}^k-\gamma_{\min}^k)L_x+2(\alpha_{\max}^k)^2L_x^2\mu_x^2\mu_\theta^2/L_\theta^4)\|x^k-x^*\|^2 \\ &+(1-\gamma_{\max}^kL_\theta^2/\mu_x+(\alpha_{\max}^k)^2C_\theta^2+2(\alpha_{\max}^k)^2\mu_x^2\mu_\theta^2/L_\theta^2)\sum_{i=1}^N\|\theta_i^k-\theta^*\|^2+(\gamma_{\max}^k)^2\nu_x^2+N(\alpha_{\max}^k)^2\nu_\theta^2 \\ &\leq \left(1-\upsilon_k\gamma_{\max}^k+\beta(\alpha_{\max}^k)^2\right)\left(\|x^k-x^*\|^2+\sum_{i=1}^N\|\theta_i^k-\theta^*\|^2\right)+\delta_k,\end{aligned}$$

where the second inequality results from invoking A13(c) through which $-\mu_\theta\alpha_{\min}^k \leq -\gamma_{\max}^kL_\theta^2/\mu_x$ and $\upsilon_k = \min\{\mu_x - 2(\gamma_{\max}^k-\gamma_{\min}^k)L_x/\gamma_{\max}^k, L_\theta^2/\mu_x\}$, $\beta = \max\{2L_x^2\mu_x^2\mu_\theta^2/L_\theta^4, C_\theta^2+2\mu_x^2\mu_\theta^2/L_\theta^2\}$, and $\delta_k = (\gamma_{\max}^k)^2\nu_x^2 +$

$N(\alpha_{\max}^k)^2\nu_\theta^2$. To show the non-summability of $(\upsilon_k\gamma_{\max}^k - \beta(\alpha_{\max}^k)^2)$, we consider two cases: (i) If $\mu_x \leq L_\theta^2/\mu_x$ then $\upsilon_k = \mu_x - 2(\gamma_{\max}^k - \gamma_{\min}^k)L_x/\gamma_{\max}^k$ and for $k > K$, $\upsilon_k \geq \mu_x - \epsilon$ where $\epsilon > 0$. Consequently, $\sum_{k>K}\upsilon_k\gamma_{\max}^k \geq \sum_{k>K}(\mu_x - \epsilon)\gamma_{\max}^k = \infty$; (ii) Alternately, if $\mu_x > L_\theta^2/\mu_x$, then for $k > K$, $\upsilon_k = L_\theta^2/\mu_x$ and $\sum_{k>K}\upsilon_k\gamma_{\max}^k = \sum_{k>K} L_\theta^2/\mu_x\gamma_{\max}^k = \infty$. Since $\alpha_{\max}^k$ is square summable from (A13), we conclude that $\sum_{k=0}^\infty(\upsilon_k\gamma_{\max}^k - \beta(\alpha_{\max}^k)^2) = \infty$. In addition, we have that

$$\lim_{k\to\infty}\frac{\delta_k}{\upsilon_k\gamma_{\max}^k - \beta(\alpha_{\max}^k)^2} = \lim_{k\to\infty}\frac{(\gamma_{\max}^k)^2\nu_x^2 + N(\alpha_{\max}^k)^2\nu_\theta^2}{\upsilon_k\gamma_{\max}^k - \beta(\alpha_{\max}^k)^2} = \lim_{k\to\infty}\frac{(\gamma_{\max}^k)\nu_x^2 + N\frac{(\alpha_{\max}^k)^2}{\gamma_{\max}^k}\nu_\theta^2}{\upsilon_k - \beta(\alpha_{\max}^k)^2/\gamma_{\max}^k} = 0,$$

where the last equality results from noting that $\lim_{k\to\infty}\gamma_{\max}^k = 0$, $\lim_{k\to\infty}(\alpha_{\max}^k)^2/\gamma_{\max}^k = 0$ and $\lim_{k\to\infty}\upsilon_k = c > 0$. Then, by invoking the super-martingale convergence theorem (Lemma 2), we have that $\|x^k - x^*\|^2 + \sum_{i=1}^N\|\theta_i^k - \theta^*\|^2 \to 0$ *a.s.* as $k \to \infty$, which implies that $x^k \to x^*$ and $\theta_i^k \to \theta^*$ *a.s.* as $k \to \infty$ for all i. ∎

A natural concern is whether the rule that relates the steplengths can be implemented in a distributed fashion without coordination. We propose a rule, first suggested by [49], in which every agent chooses a positive integer and the required coordination statement holds. We view this as a protocol that may be employed for developing distributed schemes. The next result ensures that for such a choice, the required assumptions hold [49].

Lemma 9 (Choice of steplength sequences)**.** *Let $\{\gamma_i^k\}$ and $\{\alpha_i^k\}$ be chosen such that $\gamma_i^k = \frac{1}{(k+N_i)^\alpha}$ and $\alpha_i^k = \frac{1}{(k+M_i)^\beta}$ where N_i and M_i are positive integers and $\frac{1}{2} < \beta < \alpha < 1$. Then, $\sum_{k=1}^\infty\gamma_{\min}^k = \infty$, $\sum_{k=1}^\infty(\gamma_{\max}^k)^2 < \infty$, $\sum_{k=1}^\infty(\alpha_{\max}^k)^2 < \infty$ and $\lim_{k\to\infty}\frac{\gamma_{\max}^k - \gamma_{\min}^k}{\gamma_{\max}^k} = 0$, $\lim_{k\to\infty}\frac{(\alpha_{\max}^k)^2}{\gamma_{\max}^k} = 0$, $\alpha_{\min}^k \geq \gamma_{\max}^k L_\theta^2/(\mu_x\mu_\theta)$ for sufficiently large k.*

Finally, we conclude this section with a non-asymptotic error bound that demonstrates that the joint scheme demonstrates the optimal rate of convergence of $\mathcal{O}(1/K)$ in mean-squared error.

Theorem 9. *Suppose (A10), (A11) and (A12) hold. Suppose* $\gamma_i^k = \lambda_{x,i}/k$ *and* $\alpha_i^k = \lambda_{\theta,i}/k$. *Let* $\mathbb{E}[\|F_i(x^k;\theta_i^k) + w_i^k\|^2] \leq M^2/N$ *and* $\mathbb{E}[\|\nabla_\theta g(\theta_i^k) + v_i^k\|^2] \leq M_\theta^2$ *for all* $x^k \in K$ *and* $\theta_i^k \in \Theta$. *Let* $\{x^k, \theta^k\}$ *be computed via Algorithm 4. We define* $\lambda_{x,\min} \triangleq \min_{1\leq i\leq N}\{\lambda_{x,i}\}$, $\lambda_{x,\max} \triangleq \max_{1\leq i\leq N}\{\lambda_{x,i}\}$, $\lambda_{\theta,\min} \triangleq \min_{1\leq i\leq N}\{\lambda_{\theta,i}\}$ *and* $\lambda_{\theta,\max} \triangleq \max_{1\leq i\leq N}\{\lambda_{\theta,i}\}$. *Suppose* $2\mu_\theta\lambda_{\theta,\min} > 1$ *and* $\mu_x\lambda_{x,\max} - 2(\lambda_{x,\max} - \lambda_{x,\min})L_x > 1$. *Then, the following hold after* K *iterations:*

$$\mathbb{E}[\|\theta_i^K - \theta^*\|^2] \leq \frac{Q_\theta(\lambda_\theta)}{K} \text{ and } \mathbb{E}[\|x^K - x^*\|^2] \leq \frac{Q_{x,\theta}(\lambda_x,\lambda_\theta)}{K},$$

$$\text{where } Q_\theta(\lambda_\theta) \triangleq \max\left\{\frac{\lambda_{\theta,\max}^2 M_\theta^2}{(2\mu_\theta\lambda_{\theta,\min}-1)}, \max_i \mathbb{E}[\|\theta_i^0 - \theta^*\|^2]\right\} \text{ and}$$
$$Q_{x,\theta}(\lambda_x,\lambda_\theta) \triangleq \max\left\{\frac{\lambda_{x,\max}^2 M^2 + \lambda_{x,\max}^2 L_\theta^2 N Q_\theta(\lambda_\theta)}{(\mu_x\lambda_{x,\max} - 2(\lambda_{x,\max}-\lambda_{x,\min})L_x - 1)}, \mathbb{E}[\|x^0 - x^*\|^2]\right\}.$$

Proof. Suppose $A_k \triangleq \frac{1}{2}\|x^k - x^*\|^2$ and $a_k \triangleq \mathbb{E}[A_k]$. Then, A_{k+1} may be bounded as follows by using the non-expansivity of the Euclidean projector:

$$\begin{aligned} A_{k+1} &\leq \frac{1}{2}\sum_{i=1}^N \|x_i^k - x_i^* - \gamma_i^k(F_i(x^k;\theta_i^k) + w_i^k))\|^2 \\ &= A_k + \frac{1}{2}\sum_{i=1}^N (\gamma_i^k)^2\|F_i(x^k;\theta_i^k) + w_i^k\|^2 - \sum_{i=1}^N \gamma_i^k(x_i^k - x_i^*)^T(F_i(x^k;\theta_i^k) + w_i^k). \end{aligned} \tag{3.13}$$

Note that $\mathbb{E}[(x_i^k - x_i^*)^T w_i^k] = 0$. By taking expectations on both sides of (3.13) and by invoking the bounds $\mathbb{E}[\|F_i(x^k;\theta_i^k) + w_i^k\|^2] \leq M^2/N$ and $\mathbb{E}[\|\nabla_\theta g(\theta_i^k) + v_i^k\|^2] \leq M_\theta^2$, it follows that

$$a_{k+1} \leq a_k + \frac{1}{2}(\gamma_{\max}^k)^2 M^2 - \sum_{i=1}^N \gamma_i^k \mathbb{E}[(x_i^k - x_i^*)^T F_i(x^k;\theta_i^k)]. \tag{3.14}$$

By (3.9) and (3.10), the last term (including the negative sign) in (3.14) can be bounded by

$$\begin{aligned} &-\sum_{i=1}^N \gamma_i^k \mathbb{E}[(x_i^k - x_i^*)^T(F_i(x^k;\theta_i^k) - F_i(x^k;\theta^*))] \\ &-\sum_{i=1}^N \gamma_i^k \mathbb{E}[(x_i^k - x_i^*)^T(F_i(x^k;\theta^*) - F_i(x^*;\theta^*))] - \sum_{i=1}^N \gamma_i^k \mathbb{E}[(x_i^k - x_i^*)^T F_i(x^*;\theta^*)] \\ &\leq -\sum_{i=1}^N \gamma_i^k \mathbb{E}[(x_i^k - x_i^*)^T(F_i(x^k;\theta_i^k) - F_i(x^k;\theta^*))] - \sum_{i=1}^N \gamma_i^k \mathbb{E}[(x_i^k - x_i^*)^T(F_i(x^k;\theta^*) - F_i(x^*;\theta^*))] \\ &\leq \gamma_{\max}^k \mu_x a_k + \gamma_{\max}^k L_\theta^2/(2\mu_x)\sum_{i=1}^N \mathbb{E}[\|\theta_i^k - \theta^*\|^2] - 2\gamma_{\max}^k \mu_x a_k + 2(\gamma_{\max}^k - \gamma_{\min}^k)L_x a_k. \end{aligned} \tag{3.15}$$

Combining (3.14) and (3.15), we get

$$a_{k+1} \leq (1 - \gamma^k_{\max}\mu_x + 2(\gamma^k_{\max} - \gamma^k_{\min})L_x)a_k + \frac{1}{2}(\gamma^k_{\max})^2 M^2 + \gamma^k_{\max} L^2_\theta/(2\mu_x) \sum_{i=1}^{N} \mathbb{E}[\|\theta^k_i - \theta^*\|^2].$$

Suppose $\alpha^k_i = \lambda_{\theta,i}/k$ for all i. Since the function $g(\theta)$ is strongly convex, we can use the standard rate estimate (cf. inequality (5.292) in [42]) to get the following

$$\mathbb{E}[\|\theta^k_i - \theta^*\|^2] \leq \frac{Q_\theta(\lambda_\theta)}{k}, \tag{3.16}$$

where $Q_\theta(\lambda_\theta) \triangleq \max\left\{\lambda^2_{\theta,\max} M^2_\theta (2\mu_\theta \lambda_{\theta,\min} - 1)^{-1}, \max_i \mathbb{E}[\|\theta^0_i - \theta^*\|^2]\right\}$ with $\lambda_{\theta,\min} > 1/(2\mu_\theta)$. Suppose $\gamma^k_i = \lambda_{x,i}/k$, allowing us to claim the following:

$$a_{k+1} \leq \left(1 - \frac{\mu_x \lambda_{x,\max} - 2(\lambda_{x,\max} - \lambda_{x,\min})L_x}{k}\right) a_k + \frac{\lambda^2_{x,\max}}{2k^2}\left(M^2 + \frac{L^2_\theta N Q_\theta(\lambda_\theta)}{\lambda_{x,\max}\mu_x}\right),$$

By assuming that $\mu_x \lambda_{x,\max} - 2(\lambda_{x,\max} - \lambda_{x,\min})L_x > 1$, the result follows by observing that $\mathbb{E}[\|x^k - x^*\|^2] \leq \frac{Q_{x,\theta}(\lambda_x,\lambda_\theta)}{k}$, where

$$Q_{x,\theta}(\lambda_x, \lambda_\theta) \triangleq \max\left\{\frac{\lambda^2_{x,\max} M^2 + \lambda^2_{x,\max} L^2_\theta N Q_\theta(\lambda_\theta)}{(\mu_x \lambda_{x,\max} - 2(\lambda_{x,\max} - \lambda_{x,\min})L_x - 1)}, \mathbb{E}[\|x^0 - x^*\|^2]\right\}.$$

∎

Remark: Surprisingly, misspecification does not lead to a degeneration in the rate of convergence of the mean-squared error but does lead to a worsening of the constant. In addition, the lack of consistency across steplengths leads to a further growth in this constant. In fact, if $\theta^0_i = \theta^*$ for every i, we obtain a rate close to that seen for perfectly specified stochastic Nash games.

3.3 Iterative fixed-point schemes for misspecified Nash-Cournot games

Inspired by the analysis of misspecified Nash-Cournot games [68, 84, 69, 71, 70], we develop an iterative fixed-point scheme. We introduce the problem in Section 3.3.1 and describe and analyze the algorithm in Sections 3.3.2 and 3.3.3, respectively. We conclude with an extension to nonlinear prices in Section 3.3.4.

3.3.1 Problem description, assumptions and background

We consider a Nash-Cournot game wherein $f_i(x) \triangleq c_i(x_i) - p(X; a^*, b^*)x_i$, where $X \triangleq \sum_{i=1}^{N} x_i$, x_i and $c_i(x_i)$ denotes the scalar output and cost function associated with firm i, θ^* denotes the true value of the misspecified parameter of the price function while K_i denotes the strategy set of firm i. Suppose the price function $p(X; a^*, b^*)$ is defined as

$$p(X; a^*, b^*) \triangleq (a^* - b^* X)\,, \tag{3.17}$$

Note that a^* represents the "choke price" at which demand plummets to zero, while b^* represents the price elasticity of demand. Inspired by [69, 84], we assume that either a^* or b^* is unknown and firm i's belief of this unknown parameter is denoted by θ_i. A natural extension is where both parameters are unknown and this will require two or more observations at each epoch, rather than a single observation of noisy prices.

Case 1 (Learning a^*): We assume that firms know b^* but are unaware of a^* ($\theta^* = a^*$); the ith firm harbors a belief on a^* denoted by θ_i and estimates the aggregate output X by X_i, then the ith firm's price estimate and the true noise-corrupted prices are defined as follows:

$$p(X_i; \theta_i, 0) \triangleq \theta_i - b^* X_i \text{ (\textbf{Estimate}) and } p(X; \theta^*, \xi) \triangleq (\theta^* + \xi) - b^* X. \text{ (\textbf{True price}).} \tag{3.18}$$

Case 2 (Learning b^*): Distinct from Case 1, firms know a^* and estimate b^* as θ_i ($\theta^* = b^*$) while the true price is corrupted by noise scaled by the aggregate output. Firm i's price estimate and the true prices are defined as follows:

$$p(X_i; \theta_i, 0) \triangleq a^* - \theta_i X_i \text{ (\textbf{Estimate}) and } p(X; \theta^*, \xi) \triangleq a^* - (\theta^* + \xi)X. \text{ (\textbf{True price}).} \tag{3.19}$$

The next assumption formalizes these two cases.

Assumption 14 (A14)**.** *Either (A14a) or (A14b) holds:*
(A14a) Firms know b^* but not a^* ($\theta^* = a^*$) and the price is defined by (3.18)*.*
(A14b) Firms know a^* but not b^* ($\theta^* = b^*$) and the price is defined by (3.19)*.*
Furthermore, the random variable ξ is defined by $\xi : \Lambda \to \mathbb{R}$, $(\Lambda, \mathcal{F}_\theta, \mathbb{P}_\theta)$ is the associated probability space and $\xi^1, \ldots, \xi^k$ are i.i.d. random variables with mean zero for all k.

Our assumption on costs is a special case of (A10).

Assumption 15 (A15)**.** *The cost function $c_i(x_i)$ is a convex and continuously differentiable function in*

x_i over K_i with Lipschitz continuous gradients with constant M_i. Furthermore, $K_1, \ldots, K_N, \Theta$ are closed, convex, and bounded sets.

As forwarded by [85], the notion of "common knowledge" in game theory extends beyond agents having access to information; specifically, two agents are assumed to have *common knowledge* of an event, if both agents know the event, agent 1 knows that agent 2 knows it, agent 2 knows that agent 1 knows it, agent 1 knows that agent 2 knows that agent 1 knows it and so on. We also assume that firms cannot observe aggregate output and firms employ a belief of aggregate output, relying on the knowledge of the cost functions and strategy sets of their competitors. Such a knowledge is assured through a *common knowledge* assumption. Collectively, these two assumptions are captured by (A16). An assumption often employed in games is that of *common knowledge*, whereby firms are aware of the costs functions and strategy sets of their competitors (see [57]). Formally, this assumption is given by the following:

Assumption 16 (A16)**.** *The common knowledge assumption holds with regard to $c_i(x_i)$ and K_i for $i = 1, \ldots, N$. Furthermore, aggregate output is unobservable.*

Several motivating examples exist in the literature detailing common knowledge; these include instances provided by [86] (the barbecue problem) and [87] (the department store problem), amongst others. While our results are agnostic to applications, it is worth emphasizing that such assumptions often hold when agents need to make their assets and costs public through suitable filings, such as in utility-based regulation (power, gas, water, etc.). This is often the case in regulatory settings (cf. [88, Pg. 78-79]). Common knowledge assumptions immediately hold when a game is designed [54, 55, 56] and agents can be endowed with the requisite knowledge. A select number of results will rely on boundedness of strategy sets, as specified by (A17).

Assumption 17 (A17)**.** *Suppose the estimator set Θ is a compact convex set in $\mathbb{R}_+$ given by $[\delta, \Delta]$ and $0 < \delta < \theta^* + \xi_k < \Delta$ for all k. Furthermore, suppose the sets $K_1, \ldots, K_N$ are bounded.*

3.3.2 Description and definition of algorithm

Our goal lies in developing schemes for learning equilibria and misspecified parameters. Unfortunately, since neither the aggregate output nor θ^* are observable, gradient/best-response schemes cannot be directly implemented. However, under (A16), every firm knows the cost functions and strategy sets of its competitors, allowing for computing the best response of all firms, based on an estimate of θ^* and the aggregate. By using the discrepancy between estimated and observed prices, each firm may construct improved estimates of the misspecified parameter. This model, while aligned, with that suggested by [69, 84] enjoys distinctions at

several levels; specifically, we allow for *constrained* problems with *nonlinear* cost functions with *noisy* price observations arising from possibly *nonlinear* price functions. Throughout this section, let $x_i^k = (x_{i1}^k, \cdots, x_{iN}^k)$ for $i = 1, \cdots, N$ and $X_i^k = \sum_{j=1}^k x_{ij}^k$ where x_{ij}^k denotes firm i's conjecture of firm j's output at the kth period and X_i^k denote firm i's estimate of aggregate output. Note that X_i^k is maintained as strictly positive by assuming that at least one of the strategy sets requires strictly positive output while the true aggregate X^k is given by $X^k \triangleq \sum_{j=1}^N x_{jj}^k$. The proposed algorithm relies on **simultaneous** updates of x_i^{k+1} and θ_i^{k+1}. Before proceeding, we define $\widehat{\theta}_i^{k+1}$, ϑ_i^{k+1}, and $\bar{\vartheta}_i^{k+1}$:

Definition of $\vartheta_i^k, \bar{\vartheta}_i^k$ and $\widehat{\theta}_i^k$: The variable ϑ_i^k is defined as follows:

$$\text{under } \textbf{(A14a)} : p(X^k; \theta^*, \xi^k) = (\theta^* + \xi^k) - b^* X^k, \quad \vartheta_i^k \triangleq p(X^k; \theta^*, \xi^k) + b^* X_i^k, \tag{3.20}$$

$$\text{under } \textbf{(A14b)} : p(X^k; \theta^*, \xi^k) = a^* - (\theta^* + \xi^k) X^k, \quad \vartheta_i^k \triangleq (a^* - p(X^k; \theta^*, \xi^k))/X_i^k. \tag{3.21}$$

Consequently, $\bar{\vartheta}_i^k$ after k steps is given by

$$\bar{\vartheta}_i^k = \frac{(k-1)\bar{\vartheta}_i^{k-1} + \vartheta_i^k}{k}. \tag{3.22}$$

Subsequently, we show that $\bar{\vartheta}_i^k$ is the sample average of $\theta^* + \xi^1, \ldots, \theta^* + \xi^k$ after k steps.
$\widehat{\theta}_i^{k+1}$, as a function θ_i^{k+1}, is defined as follows:

$$\widehat{\theta}_i^{k+1}(\theta_i^{k+1}) \triangleq \frac{1}{k+1}\theta_i^{k+1} + \frac{k}{k+1}\bar{\vartheta}_i^k. \tag{3.23}$$

(a) Update of $x_{i1}^{k+1}, \ldots, x_{iN}^{k+1}$: Under (A16), firm i can compute the Nash equilibrium, contingent on its choice of θ_i^{k+1}, and is a fixed-point of the best-response map:

$$x_{ij}^{k+1} \in \underset{x_j \in K_j}{\operatorname{argmin}} \left[c_j(x_{ij}^{k+1}) - p(X_i^{k+1}; \widehat{\theta}_i^{k+1}(\theta_i^{k+1}), 0) x_{ij}^{k+1} + \frac{1}{2}\epsilon^k \|x_{ij}^{k+1}\|^2 \right], \quad j = 1, \ldots, N. \tag{$\mathbf{BR}_{ij}^x(x_{i,-j}^{k+1}, \theta_i^{k+1})$}$$

(b) Update of θ_i^{k+1}: Firm i defines the difference between the price observed at the kth step $p(X^k; \theta^*, \xi^k)$ and its estimate $p(X_i^{k+1}; \widehat{\theta}_i^{k+1}, 0)$ as $\widetilde{p}_i^{k+1}(\theta_i^{k+1}, X_i^{k+1})$:

$$\widetilde{p}_i^{k+1}(\theta_i^{k+1}, X_i^{k+1}) := \begin{cases} p(X_i^{k+1}; \hat{\theta}_i^{k+1}(\theta_i^{k+1}), 0) - p(X^k; \theta^*, \xi^k), & \text{under } \textbf{(A14a)} \\ p(X^k; \theta^*, \xi^k) - p(X_i^{k+1}; \hat{\theta}_i^{k+1}(\theta_i^{k+1}), 0). & \text{under } \textbf{(A14b)} \end{cases}$$

Then suppose $t_i^{k+1}(X_i^{k+1})$ denotes a unique solution to

$$\widetilde{p}_i^{k+1}(\theta_i^{k+1}, X_i^{k+1}) + \epsilon^k \theta_i^{k+1} = 0$$

implying that

$$t_i^{k+1}(X_i^{k+1}) = \begin{cases} [(k+1)(p(X^k;\theta^*,\xi^k) + b^* X_i^{k+1}) - k\bar{\vartheta}_i^k]/(1+(k+1)\epsilon^k), & \text{under } \textbf{(A14a)} \\ [(k+1)(a^* - p(X^k;\theta^*,\xi^k)) - k\bar{\vartheta}_i^k X_i^{k+1}]/(X_i^{k+1} + (k+1)\epsilon^k), & \text{under } \textbf{(A14b)} \end{cases} \tag{3.24}$$

Suppose δ and Δ are lower and upper bounds of Θ, respectively. We can update θ_i^{k+1} as follows:

$$\theta_i^{k+1} = \begin{cases} \delta, & \text{if } \ t_i^{k+1}(X_i^{k+1}) < \delta \\ t_i^{k+1}(X_i^{k+1}), & \text{if } \ \delta \le t_i^{k+1}(X_i^{k+1}) \le \Delta \\ \Delta. & \text{if } \ t_i^{k+1}(X_i^{k+1}) > \Delta \end{cases} \qquad (\mathbf{BR}_i^{\theta}(X_i^{k+1}))$$

Algorithm 5 (**Iterative fixed-point and learning**). **Step 0.** *Given a sequence* $\{\epsilon^k\} \downarrow 0$, *and* γ_x, γ_θ. $k = 0; \sum_{j=1}^N x_{jj}^0 = X^0; p(X^0;\theta^*,\xi^0) := a^* - b^* X^0; \epsilon^0 > 0$; $\bar{\vartheta}_i^0 = 0$ *for* $i = 1, \ldots, N$.

Step 1. *For* $i = 1, \ldots, N$, *if* $X_i^{k+1} = \sum_{j=1}^N x_{ij}^{k+1}$, *then* $\{x_{i1}^{k+1}, \ldots, x_{iN}^{k+1}, \theta_i^{k+1}\}$ *is a solution to the following system:*

$$\begin{aligned} & x_{ij}^{k+1} \ \textit{solves} \ \mathbf{BR}_{ij}^x(x_{i,-j}^{k+1}, \theta_i^{k+1}), \quad j = 1, \ldots, N \\ & \theta_i^{k+1} \ \textit{solves} \ \mathbf{BR}_i^{\theta}(X_i^{k+1}). \end{aligned} \tag{3.25}$$

Step 2. *For* $i = 1, \ldots, N$, $\bar{\vartheta}_i^{k+1}$ *is updated as follows:*

$$\bar{\vartheta}_i^{k+1} = \frac{k\bar{\vartheta}_i^k + \vartheta_i^{k+1}}{k+1}. \tag{3.26}$$

Step 3. *If* $k > \bar{K}$, *stop; else* $k := k+1$ *and go to Step 1.*

3.3.3 Analysis of noise-corrupted iterative fixed-point schemes

In this subsection, we analyze our iterative fixed-point scheme and partition the discussion as follows: (i) First, we provide a brief discussion as to why the update specified by (3.25) can be succinctly captured by the solution to a **single** variational equality problem; (ii) Second, we provide a brief sketch of the results to follows; and (iii) We provide the convergence theory.

(i) Equivalence of (3.25) **to a fixed-point problem:** First, any best response of a convex optimization

problem is equivalent to a solution of a suitable variational inequality problem [20]:

$$\left[y_i^* \ \in \ \underset{y_i \in \mathcal{Y}_i}{\operatorname{argmin}} \ d_i(y_i)\right] \Leftrightarrow \ \left[y_i^* \ \text{ solves VI}(\mathcal{Y}_i, \nabla_{y_i} d_i)\right],$$

where d_i is a convex function in y_i over a convex set $\mathcal{Y}_i$. In fact, given a collection of functions $d_i(y_i; y_{-i})$ that are convex in y_i over convex sets $\mathcal{Y}_i$ for all y_{-i} with $y_{-i} \triangleq (y_j)_{i \neq j}$, the coupled best response is equivalent to the solution of a single variational inequality problem [27]:

$$\left\{ \begin{array}{c} \left[y_1^* \in \underset{y_1 \in \mathcal{Y}_1}{\operatorname{argmin}} \ d_1(y_1, y_{-1})\right] \Leftrightarrow \left[y_1^* \ \text{ solves VI}(\mathcal{Y}_1, \nabla_{y_1} d_1(\bullet, y_{-1}^*))\right] \\ \vdots \\ \left[y_N^* \in \underset{y_N \in \mathcal{Y}_N}{\operatorname{argmin}} \ d_N(y_N, y_{-N})\right] \Leftrightarrow \left[y_N^* \ \text{ solves VI}(\mathcal{Y}_N, \nabla_{y_N} d_N(\bullet, y_{-N}^*))\right] \end{array} \right\} \Leftrightarrow y^* \text{ solves VI}(\mathcal{Y}, F),$$

where $\mathcal{Y} \triangleq \prod_{i=1}^N \mathcal{Y}_i$ and $F(y) = (\nabla_{y_i} d_i(y_i, y_{-i}))_{i=1}^N$. Finally, any solution to a variational inequality problem is a fixed point of a suitably defined problem where γ is a positive scalar:

$$[y^* \text{ solves VI}(\mathcal{Y}, F)] \Leftrightarrow \left[y^* = \Pi_{\mathcal{Y}}(y^* - \gamma F(y^*))\right].$$

By using this avenue, the problem $(\mathbf{BR}_{ij}^x(x_{i,-j}^{k+1}, \theta_i^{k+1}))$ is the set of coupled fixed-point problems:

$$x_{ij}^{k+1} = \Pi_{K_j}\left(x_{ij}^{k+1} - \gamma\left(\nabla_{x_{ij}} f_j(x_i^{k+1}; \widehat{\theta}_i^{k+1}(\theta_i^{k+1})) + \epsilon^k x_{ij}^{k+1}\right)\right), \ j = 1, \ldots, N, \tag{3.27}$$

where $f_j(x_i^{k+1}; \widehat{\theta}_i^{k+1}(\theta_i^{k+1})) = c_j(x_{ij}^{k+1}) - p(X_i^{k+1}; \widehat{\theta}_i^{k+1}(\theta_i^{k+1}))x_{ij}^{k+1}$. Similarly, $(\mathbf{BR}_i^\theta(X_i^{k+1}))$ can be stated as the following fixed-point problem:

$$\theta_i^{k+1} = \Pi_\Theta\left(\theta_i^{k+1} - \gamma\left(\widetilde{p}_i^{k+1}(\theta_i^{k+1}, X_i^{k+1}) + \epsilon^k \theta_i^{k+1}\right)\right). \tag{3.28}$$

Before proceeding, we shed some light on this equivalence. Suppose the root of $\widetilde{p}_i^{k+1}(\theta_i^{k+1}, X_i^{k+1}) + \epsilon^k \theta_i^{k+1} = 0$ is denoted by t_i^{k+1}. Then from (3.28), this implies that $t_i^{k+1} = \Pi_\Theta\left(t_i^{k+1}\right)$. Consequently, if $t_i^{k+1} \in \Theta \triangleq [\delta, \Delta]$, then $\theta_i^{k+1} = t_i^{k+1}$ while $\theta_i^{k+1} = \delta($ or $\Delta)$, if $t_i^{k+1} < \delta($ or $> \Delta)$. But this is equivalent to $(\mathbf{BR}_i^\theta(X_i^{k+1}))$.

We define $z_i^{k+1} \triangleq (x_{i1}^{k+1}, \ldots, x_{iN}^{k+1}, \theta_i^{k+1})$. Then, z_i^{k+1} solves the coupled fixed-point problem (3.27) –

(3.28) if and only if z_i^{k+1} solves $\text{VI}(\mathcal{Z}, F^{k+1})$ where

$$\mathcal{Z} \triangleq \prod_{i=1}^{N} K_i \times \Theta \text{ and } F^{k+1}(z_i^{k+1}) = \begin{pmatrix} \nabla_{x_{i1}} f_1(x_i^{k+1}; \widehat{\theta}_i^{k+1}(\theta_i^{k+1})) \\ \vdots \\ \nabla_{x_{iN}} f_N(x_i^{k+1}; \widehat{\theta}_i^{k+1}(\theta_i^{k+1})) \\ \widetilde{p}_i^{k+1}(\theta_i^{k+1}, X_i^{k+1}) \end{pmatrix} + \epsilon^k z_i^{k+1}.$$

In sum, the coupled best response scheme (3.25) is equivalent to the coupled fixed-point problem (3.27) – (3.28), which is also equivalent to the variational inequality problem $\text{VI}(\mathcal{Z}, F^{k+1})$.

(ii) Sketch of results: We first show that the coupled best response scheme given by (3.25) always admits a unique solution (Prop. 6). Theorem 10 shows that the sequence $\{x_i^k, \widehat{\theta}_i^k\} \to \{x^*, \theta^*\}$ as $k \to \infty$ in an a.s. sense. This proof relies on showing that $\widehat{\theta}_i^k \to \theta^*$ as $k \to \infty$ in an a.s. sense. Then, if the solution $x_i^{k+1}(\widehat{\theta}_i^{k+1})$ is a continuous function in $\widehat{\theta}_i^{k+1}$ (Prop. 8), we may conclude that $\lim_{k\to\infty} x_i^{k+1}(\widehat{\theta}_i^{k+1}) = x_i^{k+1}(\theta^*) = x^*$, where the last equality follows from noting that

Proposition 6. *Suppose (A14), (A15) and (A16) hold. If $k \geq 0$ and $\epsilon^k > 0$, and given $p(X^k; \theta^*, \xi^k)$ and $\{\bar{\vartheta}_i^k\}_{i=1}^N$, the following hold:*

(a) Under (A14a), the solution to (3.25) *is a singleton.*

(b) Under (A14b), the solution to (3.25) *is a singleton.*

Proof. It suffices to show that given $p(X^k; \theta^*, \xi^k)$ and $\{\bar{\vartheta}_i^k\}_{i=1}^N$, the variational inequality $\text{VI}(\mathcal{Z}, F^{k+1})$ has a unique solution for each i. Now, for simplicity, we ignore the superscript k for all variables. Given p, $\bar{\vartheta}_i$, i and k, let $H(z_i)$ denote the Jacobian matrix $\nabla F(z_i)$ of F at $z_i \in \mathcal{Z}$. We will proceed to show that $H(z_i)$ is a **P**-matrix for all $z_i \in \widetilde{\mathcal{Z}}$ in part (a) and a $\mathbf{P_0}$-matrix for all $z_i \in \widetilde{\mathcal{Z}}$ in part (b) where $\mathcal{Z} \subset \widetilde{\mathcal{Z}}$ and $\widetilde{\mathcal{Z}}$ is a rectangle. Then, by invoking Proposition 3.5.9 in [20], the associated mapping F is **P**-mapping on $\widetilde{\mathcal{Z}}$ in part (a) and a $\mathbf{P}_0$-mapping on $\widetilde{\mathcal{Z}}$ in part (b). Consequently, by Theorem 3.5.15 in [20], the regularized variational inequality $\text{VI}(\mathcal{Z}, F^{k+1})$ has a unique solution in both parts (a) and (b). Specifically, we employ a rectangular $\widetilde{\mathcal{Z}}$ defined as $\widetilde{\mathcal{Z}} \triangleq [0, \infty)^N \times \Theta$, where Θ is a compact set in $(0, \infty)$. **(a)** Given $z_i \in \widetilde{\mathcal{Z}}$, let H_i denote $H(z_i)$. Then,

$$H_i = \begin{pmatrix} A_i & B \\ C & D \end{pmatrix}, \tag{3.29}$$

where $A_i = b^*(I + ee^T) + E_i$, $B = -\frac{1}{k+1}e$, $C = -b^*e^T$, $D = \frac{1}{k+1}$, e denotes the column of ones in $\mathbb{R}^N$, E_i is

an $N \times N$ diagonal matrix with $c_j''(x_{ij})$ as its jth diagonal entry. Since, the nonnegativity of $c_j''(x_{ij})$ follows from the convexity of costs, E_i is a nonnegative diagonal matrix and is therefore positive semidefinite. Recall that the sum of a diagonal positive semidefinite matrix and a **P**-matrix is a **P**-matrix and it suffices to show that H_i is a **P**-matrix when $E_i = 0$. This amounts to showing that the principal minors of H are positive.

Since A_i and D are **P**-matrices, we only consider the principal submatrix H_α of H_i, where $\alpha \subseteq \{1, \ldots, N\}$ is a nonempty index set and H_α is given by $H_\alpha \triangleq \begin{pmatrix} A_\alpha & B_\alpha \\ C_\alpha & D \end{pmatrix}$, where $A_\alpha = b^*(I_{n_\alpha} + e^{n_\alpha}(e^{n_\alpha})^T), B_\alpha = -\frac{1}{k+1}e^{n_\alpha}, C_\alpha = -b^*(e^{n_\alpha})^T$, and I_{n_α} and e^{n_α} denote the identity matrix and the column of ones in $\mathbb{R}^{n_\alpha \times n_\alpha}$ and $\mathbb{R}^{n_\alpha}$, respectively, with $n_\alpha = |\alpha|$. Since $A_\alpha^{-1} = \frac{1}{b^*}\left(I_{n_\alpha} - \frac{1}{n_\alpha+1}e^{n_\alpha}(e^{n_\alpha})^T\right)$, we have

$$\begin{aligned} C_\alpha A_\alpha^{-1} B_\alpha &= \frac{1}{k+1}(e^{n_\alpha})^T \left(I_{n_\alpha} - \frac{1}{n_\alpha+1}e^{n_\alpha}(e^{n_\alpha})^T\right) e^{n_\alpha} \\ &= \frac{1}{k+1}\left(n_\alpha - \frac{n_\alpha^2}{n_\alpha+1}\right) = \frac{1}{k+1}\left(\frac{n_\alpha}{n_\alpha+1}\right). \end{aligned}$$

It follows that $D - C_\alpha A_\alpha^{-1} B_\alpha = \frac{1}{k+1} - \frac{1}{k+1}\left(\frac{n_\alpha}{n_\alpha+1}\right) = \frac{1}{k+1}\left(\frac{1}{n_\alpha+1}\right) > 0$. Since $\det(A_\alpha) > 0$, we have $\det(H_\alpha) = \det(A_\alpha)\det\left(D - C_\alpha A_\alpha^{-1} B_\alpha\right) > 0$ for all $\alpha \subseteq \{1, \ldots, N\}$ with $\alpha \neq \emptyset$. Therefore, H is a **P**-matrix.
(b) Analogous to our approach for (a), we consider a matrix H_i, given by $H_i = \nabla F(z_i)$. Then,

$$H_i = \begin{pmatrix} A_i & B_i \\ C_i & D_i \end{pmatrix}, \tag{3.30}$$

where $A_i = \hat{b}_i(I + ee^T) + E_i$, $B_i = \frac{1}{k+1}(x_i + (e^T x_i)e)$, $C_i = \hat{b}_i e^T$, and $D_i = \frac{1}{k+1}(e^T x_i)$, where $\hat{b}_i = \frac{1}{k+1}b_i + \frac{k}{k+1}\bar{b}_i$, $x_i = (x_{i1}, \ldots, x_{iN})^T$, e denotes the column of ones in $\mathbb{R}^N$, and E_i is an $N \times N$ diagonal matrix with $c_j''(x_{ij})$ as its jth diagonal entry. Recall that the sum of a diagonal positive semidefinite matrix and a $\mathbf{P}_0$-matrix is a $\mathbf{P}_0$-matrix. As in (a), it suffices to show that H is a $\mathbf{P}_0$-matrix when $E_i = 0$.

Since A_i and D_i are $\mathbf{P}_0$-matrices, we restrict our attention to the principal submatrix H_α of H_i, where $\alpha \subseteq \{1, \ldots, N\}$ is a nonempty index set, and H_α is given by $H_\alpha \triangleq \begin{pmatrix} A_\alpha & B_\alpha \\ C_\alpha & D_i \end{pmatrix}$, where $A_\alpha = \hat{b}_i(I_{n_\alpha} + e^{n_\alpha}(e^{n_\alpha})^T), B_\alpha = \frac{1}{k+1}(x_\alpha + (e^T x_i)e^{n_\alpha}), C_\alpha = \hat{b}_i(e^{n_\alpha})^T$, and I_{n_α} and e^{n_α} denote the identity matrix and the column of ones in $\mathbb{R}^{n_\alpha \times n_\alpha}$ and $\mathbb{R}^{n_\alpha}$, respectively, with $n_\alpha = |\alpha|$. Then, the following hold:

(1) If $\hat{b}_i = 0$, then $A_\alpha = 0$ and $C_\alpha = 0$, which implies $\det(H_\alpha) = 0$.

(2) If $\hat{b}_i > 0$, then $A_\alpha^{-1} = \frac{1}{\hat{b}_i}(I_{n_\alpha} - \frac{1}{n_\alpha+1}e^{n_\alpha}(e^{n_\alpha})^T)$. So, we have

$$\begin{aligned} C_\alpha A_\alpha^{-1} B_\alpha &= \frac{1}{k+1}(e^{n_\alpha})^T \left(I_{n_\alpha} - \frac{1}{n_\alpha+1}e^{n_\alpha}(e^{n_\alpha})^T\right)(x_\alpha + (e^T x_i)e^{n_\alpha}) \\ &= \frac{1}{k+1}((e^{n_\alpha})^T - \frac{n_\alpha}{n_\alpha+1}\left(e^{n_\alpha}\right)^T)(x_\alpha + (e^T x_i)e^{n_\alpha}) \\ &= \frac{1}{(k+1)(n_\alpha+1)}\left((e^{n_\alpha})^T x_\alpha + n_\alpha(e^T x_i)\right). \end{aligned}$$

$$\begin{aligned} \implies D_i - C_\alpha A_\alpha^{-1} B_\alpha &= \frac{1}{k+1}e^T x_i - \frac{1}{(k+1)(n_\alpha+1)}\left((e^{n_\alpha})^T x_\alpha + n_\alpha(e^T x_i)\right) \\ &= \frac{1}{(k+1)(n_\alpha+1)}\left(e^T x_i - (e^{n_\alpha})^T x_\alpha\right) \geq 0. \end{aligned}$$

Since $\det(A_\alpha) > 0$, we have $\det(H_\alpha) = \det(A_\alpha)\det\left(D_i - C_\alpha A_\alpha^{-1} B_\alpha\right) \geq 0$.

Therefore, $\det(H_\alpha) \geq 0$ for all nonempty $\alpha \subseteq \{1, \ldots, N\}$, implying that H_i is a $\mathbf{P_0}$-matrix. ∎

Having shown that the coupled best response scheme has a unique solution, we proceed to show a Lipschitzian property on the solution set of (3.25) with respect to the parameter θ (Prop. 8). Before that, we provide some preliminary results. The strong monotonicity and Lipschitz continuity of the mapping $F(x)$ can be easily shown under (A15).

Lemma 10. *Consider the mapping $F(x)$ defined by* (3.1) *and suppose (A15) holds. Then $F(x)$ is a strongly monotone Lipschitz continuous mapping.*

Proof. Let $g(x) = (c_1'(x_1), \ldots, c_N'(x_N))^T$ and $e = (1, \ldots, 1)^T$. Then, we have $F(x) = g(x) + b^*(x + Xe) - a^*e$, where $X = \sum_{i=1}^N x_i$. Note that $g(x)$ is monotone in x. Thus, we have for $x, y \in K$

$$\begin{aligned} (F(x) - F(y))^T(x-y) &= (g(x) - g(y))^T(x-y) + b^*(x-y)^T(x-y) + b^*(X-Y)e^T(x-y) \\ &\geq b^*(x-y)^T(x-y) + b^*(X-Y)^T(X-Y) \geq b^*\|x-y\|^2. \end{aligned}$$

This implies that $F(x)$ is strongly monotone in x with constant b^*. Note that $g(x)$ is Lipschitz continuous on K with constant M, where $M \triangleq \max_i\{M_i\}$. The Lipschitz continuity of $F(x)$ is easily shown:

$$\begin{aligned} \|F(x) - F(y)\| &= \|g(x) - g(y)\| + b^*\|x-y\| + b^*\|(X-Y)e\| \\ &\leq M\|x-y\| + b^*\|x-y\| + b^*\|ee^T\|\|x-y\| = L\|x-y\|, \end{aligned}$$

where $L = M + b^* + b^*\|ee^T\|$. It follows that $F(x)$ is Lipschitz continuous with constant L. ∎

This allows for claiming the existence and uniqueness of a Nash-Cournot equilibrium when the price function is affine.

Proposition 7. *Consider a Nash-Cournot game in which the ith player solves (Opt(x_{-i})) and the price is determined by* (3.17)*. Furthermore, suppose (A15) holds. Then, the associated Nash-Cournot game admits a unique equilibrium.*

Proof. From Lemma 10, the associated variational inequality VI(K, F) has a strongly monotone mapping $F(x)$ over K. Consequently, VI(K, F) admits a unique solution [20]. ∎

Now, we state the Lipschitzian property on the solution set of (3.25). This proof is inspired by a related result presented by [89].

Proposition 8. *Consider a VI$(K, F(\bullet;\theta))$ where $F(x;\theta)$ is strongly monotone in x over K for all $\theta \in \Theta$, Lipschitz continuous in x for all $\theta \in \Theta$ and Lipschitz continuous in θ for all $x \in K$. Then, the following hold: (a) If $x(\theta)$ denotes the solution of VI$(K, F(\bullet;\theta))$, then $x(\theta)$ is Lipschitz continuous in θ for all $\theta \in \Theta$. (b) Given an $\epsilon > 0$, if $x(\theta, \epsilon)$ denotes the solution of VI$(K, F(\bullet;\theta) + \epsilon\mathbf{I})$, then $x(\theta, \epsilon)$ is Lipschitz continuous in θ and ϵ.*

Proof. Consider $\theta_1, \theta_2 \in \Theta$ and let $F_i(\cdot) := F(\cdot, \theta_i)$, $i = 1, 2$. Let x_i be a solution of VI(K, F_i) for $i = 1, 2$. By the assumption of strong monotonicity on the map, we have that

$$(x_1 - x_2)^T(F_1(x_1) - F_1(x_2)) \geq c\|x_2 - x_1\|^2, \tag{3.31}$$

for some constant $c > 0$ (assumed to be independent of θ_1). Since x_1 is a solution of VI(K, F_1), it follows that $(x_2 - x_1)^T F_1(x_1) \geq 0$, which together with (3.31) implies

$$(x_2 - x_1)^T F_1(x_2) \geq c\|x_2 - x_1\|^2. \tag{3.32}$$

We may express (3.32) as $(x_2 - x_1)^T(F_1(x_2) - F_2(x_2) + F_2(x_2)) \geq c\|x_2 - x_1\|^2$. Now since x_2 is the solution of VI(K, F_2), it follows that $(x_2 - x_1)^T F_2(x_2) \leq 0$. Consequently we obtain

$$\|x_2 - x_1\|\|F_1(x_2) - F_2(x_2)\| \geq (x_2 - x_1)^T(F_1(x_2) - F_2(x_2)) \geq c\|x_2 - x_1\|^2. \tag{3.33}$$

By Lipschitz continuity of $F(x, \theta)$ (assuming it is uniform in x), we have that $\|F_1(x_2, \theta_1) - F_2(x_2, \theta_2)\| \leq L_\theta\|\theta_2 - \theta_1\|$, and hence by (3.33) $L_\theta\|x_2 - x_1\|\|\theta_2 - \theta_1\| \geq c\|x_2 - x_1\|^2$. It follows that $\|x_2 - x_1\| \leq L_\theta c^{-1}\|\theta_2 - \theta_1\|$.

To show (b), let $x(\theta_i, \epsilon_j)$ be the solution of VI$(K, G_{ij}(\cdot))$, where $G_{ij}(\cdot) = F(\cdot, \theta_i) + \epsilon_j\mathbf{I}$. We begin by applying the triangle inequality to obtain that $\|x(\theta_1, \epsilon_1) - x(\theta_2, \epsilon_2)\| \leq \|x(\theta_1, \epsilon_1) - x(\theta_2, \epsilon_1)\| + \|x(\theta_2, \epsilon_1) -$

$x(\theta_2, \epsilon_2)\|$. Since G_{i1} is strongly monotone in x with constant $c + \epsilon_1$ and Lipschitz continuous in θ with constant L_θ, respectively, we have that the first term is bounded by $L_\theta(c+\epsilon_1)^{-1}\|\theta_2 - \theta_1\|$ as a result from part (a). Before proceeding, the Lipschitz continuity of $F(x;\theta) + \epsilon I$ with respect to ϵ can be obtained as

$$\|(F(x;\theta) + \epsilon_2 x) - (F(x;\theta) + \epsilon_1 x)\| \leq \|x\|\|\epsilon_1 - \epsilon_2\| \leq D\|\epsilon_1 - \epsilon_2\|.$$

Since G_{2j} is strongly monotone in x with constant $c + \epsilon_j$ and Lipschitz continuous in ϵ with constant D, respectively, we have that the second term is bounded by $D(c+\epsilon_1)^{-1}\|\epsilon_2 - \epsilon_1\|$ as a result from part (a). Consequently, we obtain that

$$\|x(\theta_1, \epsilon_1) - x(\theta_2, \epsilon_2)\| \leq L_\theta(c+\epsilon_1)^{-1}\|\theta_2 - \theta_1\| + D(c+\epsilon_1)^{-1}\|\epsilon_2 - \epsilon_1\|.$$

The Lipschitz continuity of $x(\theta, \epsilon)$ with respect to its parameters follows. ∎

Notice that the solution x_i^{k+1} to the problem $(\mathbf{BR}_{ij}^x(x_{i,-j}^{k+1}, \theta_i^{k+1}))$ is the solution to the variational inequality problem $\text{VI}(\prod_{i=1}^N K_i, F_x^{k+1}(\bullet; \widehat{\theta}_i^{k+1}) + \epsilon^k \mathbf{I})$ where $F_x^{k+1}(x_i^{k+1}; \widehat{\theta}_i^{k+1}) = \left(\nabla_{x_{ij}} f_j(x_i^{k+1}; \widehat{\theta}_i^{k+1})\right)_{j=1}^N$. Based on Lemma 10 and Prop. 8, $x_i^{k+1} = x_i^{k+1}(\widehat{\theta}_i^{k+1}, \epsilon^k)$ is a continuous function of $(\widehat{\theta}_i^{k+1}, \epsilon^k)$.

We may now show that the iterative fixed-point scheme produces a sequence of iterates that converge almost surely to the true equilibrium and allow for learning the true parameter.

Theorem 10 (Global convergence of iterative fixed-point scheme)**.** *Suppose (A14), (A15), (A16) and (A17) hold. Let $\{x_i^k, \hat{\theta}_i^k\}$ be computed via Algorithm 5 for $i = 1, \ldots, N$. Then $\hat{\theta}_i^k \to \theta^*$ and $x_i^k \to x^*$ almost surely for $i = 1, \ldots, N$, where x^* is a solution of the variational inequality* (3.2).

Proof. Suppose $k \geq 0$. At the kth iteration, $\tilde{p}_i^k$ is a function of $\hat{\theta}_i^{k+1}$, which is a function of $\bar{\vartheta}_i^k$. Consequently, the fixed-point problem (3.25) is a function of $\bar{\vartheta}_i^k$; at the outset, $\bar{\vartheta}_i^0$ is zero for $i = 1, \ldots, N$ and every agent is faced by (3.25) with the same parametrization. Since (3.25) has a unique solution (Prop. 6), it follows that $x_{i,\bullet} = x_{j,\bullet}$ for $i \neq j$ and $x_{ij}^k = x_{jj}^k$. Therefore, Given $p(X^k; \theta^*, \xi^k)$ and $\{\bar{\vartheta}_i^k\}_{i=1}^N$,the solution $(x_i^{k+1}, \theta_i^{k+1})$ to (3.25) satisfies $x_{ij}^{k+1} = x_{jj}^{k+1}$ for all i, j. Thus, for all $k \geq 0$ and all i, we have that

$$p(X^k; \theta^*, \xi^k) = \begin{cases} (a^* + \xi^k) - b^* \sum_{j=1}^N x_{jj}^k = (a^* + \xi^k) - b^* \sum_{j=1}^N x_{ij}^k, & \text{under } \textbf{(A14a)}, \\ a^* - (\theta^* + \xi^k) \sum_{j=1}^N x_{jj}^k = a^* - (\theta^* + \xi^k) \sum_{j=1}^N x_{ij}^k, & \text{under } \textbf{(A14b)}. \end{cases}$$

Since for all $k \geq 0$ and all i,

$$\vartheta_i^k = \begin{cases} p(X^k; \theta^*, \xi^k) + b^* X_i^k, & \text{under } \textbf{(A14a)}, \\ (a^* - p(X^k; \theta^*, \xi^k))/X_i^k, & \text{under } \textbf{(A14b)}. \end{cases}$$

we have $\vartheta_i^k = \theta^* + \xi^k$ for all i. As a result, after k iterative fixed-point steps, we obtain k samples $\{\theta^* + \xi^1, \ldots, \theta^* + \xi^k\}$ of the estimated parameter. Since for all $k \geq 0$ and all i, $\bar{\vartheta}_i^k = \frac{(k-1)\bar{\vartheta}_i^{k-1} + \vartheta_i^k}{k}$, the sample mean of the estimated parameter is given by $\bar{\vartheta}_i^k$, i.e., $\bar{\vartheta}_i^k = \frac{\sum_{l=1}^k (\theta^* + \xi^l)}{k}$. Therefore, $\bar{\vartheta}_i^k \to \theta^*$ *a.s.* as $k \to \infty$, which implies by the boundedness of $\{\theta_i^k\}$ that for all i

$$\widehat{\theta}_i^{k+1} = \frac{1}{k+1}\theta_i^{k+1} + \frac{k}{k+1}\bar{\vartheta}_i^k \to \theta^* \quad a.s. \text{ as } k \to \infty,$$

by the strong law of large numbers. By Proposition 8, $x_i^{k+1} = x_i^{k+1}(\widehat{\theta}_i^{k+1}, \epsilon^k)$ is a continuous function of $(\widehat{\theta}_i^{k+1}, \epsilon^k)$, and $x_i^{k+1}(\theta^*, 0) = x^*$. Therefore, $x_i^{k+1} \to x^*$ *a.s.* as $k \to \infty$. ∎

Remark: We emphasize that this scheme requires each agent to effectively solve a suitably defined variational inequality problem, similar to the centralized problem seen in [69, 84]. Such schemes more closely tied to best-response schemes than the gradient-based approaches presented in the previous section. Yet, it is worth emphasizing that the computational complexity of the best-response step is of the same order as that of solving a strongly convex optimization problem, which is the problem that arises in computing a projected gradient step [27].

3.3.4 Extension to nonlinear price functions

We now consider a generalization to nonlinear prices defined as follows:

$$p(X; \theta^*, \xi) \triangleq \begin{cases} a^* - b^* X^\sigma + \xi, \\ a^* - (b^* + \xi) X^\sigma. \end{cases} \tag{3.34}$$

This nonlinear price function has been examined by [49] where a discussion of the strict monotonicity of the associated mapping is presented (Lemma 11(a)). Specifically, the equilibrium of the Nash-Cournot game are captured by VI(K, F) where $F(x)$ is defined as

$$F(x) \triangleq \left(c_i'(x_i) - (a^* - b^* X^\sigma) + \sigma b^* X^{\sigma-1} x_i \right)_{i=1}^N. \tag{3.35}$$

In the next result, the mapping $F(x)$ is strongly monotone for all $x \in K$ if $\nabla F(x)$ is a diagonally dominant matrix for all $x \in K$.

Lemma 11. *Consider the mapping $F(x)$ defined in (3.35). Suppose (A15) holds, $N < \frac{3\sigma-1}{\sigma-1}$ and $\sigma > 1$. Then the following hold:*

(a) $F(x)$ is a strictly monotone mapping over K;

(b) Suppose $X \geq \eta$ for some $\eta > 0$, then $F(x)$ is a strongly monotone mapping over K.

Proof. (a) Strict monotonicity of $F(x)$ is implied by the positive definiteness of the Jacobian $\nabla F(x)$. This is given by $\nabla F(x) = J_1 + J_2 + J_3$, where $J_2 = 2b^*\sigma X^{\sigma-1}ee^T$ and

$$J_1 = \begin{pmatrix} c_1''(x_1) & & \\ & \ddots & \\ & & c_N''(x_N) \end{pmatrix} \text{ and }, J_3 = b^*\sigma(\sigma-1)X^{\sigma-2} \begin{pmatrix} \frac{X}{\sigma-1} + x_1 & \dots & x_1 \\ \vdots & \ddots & \vdots \\ x_N & \dots & \frac{X}{\sigma-1} + x_N \end{pmatrix}.$$

Since $c_i(x_i)$ is a convex function in x_i for all i, J_1 is a positive semidefinite matrix. J_2, compactly stated as $2b^*\sigma X^{\sigma-1}ee^T$, is also a positive semidefinite matrix. As a consequence, positive definiteness of $\nabla F(x)$ follows from the diagonal dominance of the following matrix:

$$b^*\sigma(\sigma-1)X^{\sigma-2} \begin{pmatrix} \frac{X}{\sigma-1} + x_1 & \dots & \frac{1}{2}(x_1 + x_N) \\ \vdots & \ddots & \vdots \\ \frac{1}{2}(x_N + x_1) & \dots & \frac{X}{\sigma-1} + x_N \end{pmatrix}.$$

By a minor rearrangement, it suffices to show the diagonal dominance of the following:

$$b^*\sigma(\sigma-1)X^{\sigma-2} \begin{pmatrix} \frac{X_{-1}}{\sigma-1} + (1 + \frac{1}{(\sigma-1)})x_1 & \dots & \frac{1}{2}(x_1 + x_N) \\ \vdots & \ddots & \vdots \\ \frac{1}{2}(x_N + x_1) & \dots & \frac{X_{-N}}{\sigma-1} + (1 + \frac{1}{(\sigma-1)})x_N \end{pmatrix},$$

where $X_{-j} \triangleq \sum_{i \neq j} x_i$. The result follows by noting that

$$\left(1 + \frac{1}{(\sigma-1)}\right) > \frac{(N-1)}{2} \text{ or } \frac{2\sigma}{\sigma-1} > N-1 \text{ or } N < \frac{3\sigma-1}{\sigma-1}.$$

(b) For $x, y \in K$, $(x-y)^T(F(x) - F(y)) = \int_0^1 (x-y)^T \nabla F(y + \alpha(x-y))(x-y)d\alpha$. Let $\tilde{x} = y + \alpha(x-y)$ an $\tilde{X} = \sum_{i=1}^N \tilde{x}_i$. Akin to $\nabla F(x)$, $\nabla F(y + \alpha(x-y)) = \tilde{J}_1 + \tilde{J}_2 + \tilde{J}_3$, where $\tilde{J}_1$ and $\tilde{J}_2$ are positive semidefinite,

and $\tilde{J}_3 = b^*\sigma(\sigma-1)\tilde{X}^{\sigma-2}\tilde{J}_4$, where

$$\tilde{J}_4 = \begin{pmatrix} \frac{\tilde{X}_{-1}}{\sigma-1} + (1+\frac{1}{(\sigma-1)})\tilde{x}_1 & \dots & \frac{1}{2}(\tilde{x}_1+\tilde{x}_N) \\ \vdots & \ddots & \vdots \\ \frac{1}{2}(\tilde{x}_N+\tilde{x}_1) & \dots & \frac{\tilde{X}_{-N}}{\sigma-1} + (1+\frac{1}{(\sigma-1)})\tilde{x}_N \end{pmatrix}$$

$$= \begin{pmatrix} \frac{\tilde{X}_{-1}}{\sigma-1} + \frac{N-1}{2}\tilde{x}_1 & \dots & \frac{1}{2}(\tilde{x}_1+\tilde{x}_N) \\ \vdots & \ddots & \vdots \\ \frac{1}{2}(\tilde{x}_N+\tilde{x}_1) & \dots & \frac{\tilde{X}_{-N}}{\sigma-1} + \frac{N-1}{2}\tilde{x}_N \end{pmatrix} + \left(\frac{\sigma}{\sigma-1} - \frac{N-1}{2}\right) I_N \triangleq \tilde{J}_5 + \rho I_N,$$

where $\tilde{J}_5$ is a positive semidefinite matrix and $\rho = (1 + \frac{1}{\sigma-1} - \frac{N-1}{2}) > 0$. Therefore,

$$\begin{aligned}(x-y)^T(F(x)-F(y)) &\geq \int_0^1 (x-y)^T\tilde{J}_3(x-y)d\alpha \\ &\geq b^*\sigma(\sigma-1)\int_0^1 (x-y)^T\tilde{X}^{\sigma-2}\tilde{J}_4(x-y)d\alpha \\ &\geq b^*\sigma(\sigma-1)\eta^{\sigma-2}\int_0^1 (x-y)^T(\tilde{J}_5+\rho I_N)(x-y)d\alpha \\ &\geq b^*\sigma(\sigma-1)\eta^{\sigma-2}\rho\|x-y\|^2,\end{aligned}$$

implying the strong monotonicity of F. ∎

Directly deriving a Lipschitzian statement on $F(x;\theta)$ in terms of θ is not easy when the price function has the prescribed nonlinear form; instead, by noting that $\nabla F(x)$ is bounded when x is bounded, allows for proving such a statement. Next, we provide a corollary of Proposition 8 where such a property is derived.

Corollary 3. *Consider a VI$(K, F(\bullet;\theta))$ where $F(x;\theta)$ is strongly monotone in x over K for all $\theta \in \Theta$, and Lipschitz continuous in θ for all $x \in K$. Also, there is a constant $R > 0$, such that $\|\nabla F(x;\theta)\| \leq R$ for all $x \in K$ and $\theta \in \Theta$. Given an $\epsilon > 0$, if $x(\theta,\epsilon)$ denotes the solution of VI$(K, F(\bullet;\theta) + \epsilon\mathbf{I})$, then $x(\theta,\epsilon)$ is Lipschitz continuous in θ and ϵ.*

Proof. By Proposition 8, it suffices to show that $F(x;\theta)$ is Lipschitz continuous in x for all $\theta \in \Theta$. For $\theta \in \Theta$, and $x, y \in K$, we have that

$$\begin{aligned}\|F(x;\theta)-F(y;\theta)\| &= \left\|\int_0^1 \nabla F(y+\alpha(x-y);\theta)(x-y)d\alpha\right\| \qquad \text{for some } \alpha \in (0,1) \\ &\leq \int_0^1 \|\nabla F(y+\alpha(x-y);\theta)\|\,\|x-y\|d\alpha \leq \int_0^1 r\|x-y\|d\alpha = r\|x-y\|,\end{aligned}$$

which implies the Lipschitz continuity in x of the mapping F. ∎

Proposition 9. *Suppose (A14a) holds. Consider the mapping $F(x)$ defined in* (3.35) *and suppose (A15) and (A17) hold. Suppose $X \geq \eta$ for some $\eta > 0$ and all $x \in K$, where $X = \sum_{i=1}^{N} x_i$. If $N < \frac{3\sigma-1}{\sigma-1}$ and $\sigma > 1$, then the following hold:*

(a) If $x(\theta)$ denotes the solution of VI$(K, F(.;\theta))$, then $x(\theta)$ is Lipschitz continuous in θ for all $\theta \in \Theta$.

(b) Given an $\epsilon > 0$, if $x(\theta, \epsilon)$ denotes the solution of VI$(K, F(.;\theta)+\epsilon\mathbf{I})$, then $x(\theta, \epsilon)$ is Lipschitz continuous in θ and ϵ.

Proof. By Lemma 11, $F(x;\theta)$ is a strongly monotone mapping over K for all $\theta \in \Theta$. By definition of F, $F(x;\theta)$ is Lipschitz continuous in θ for all $x \in K$. By definition of ∇F and boundedness of $x \in K$, $\nabla F(x;\theta)$ is bounded for $x \in K$ and $\theta \in \Theta$. Then, the conclusion follows from Corollary 3. ∎

We may now show that the fixed-point problem yields a unique solution.

Proposition 10. *Suppose (A15) and (A16) hold. Let the price be given by* (3.34). *If $N < \frac{3\sigma-1}{\sigma-1}$ and $\sigma > 1$, then given $p^k(\xi^k)$ and $\{\bar{\theta}_i^k\}_{i=1}^N$, the solution to* (3.25) *is a singleton.*

Proof. Given p, $\bar{\theta}_i$, i and k, let $H(z_i)$ denote the Jacobian matrix $\nabla F(z_i)$ of the mapping F at $z_i \in \widetilde{\mathcal{Z}}$. Then, as in Proposition 6, it suffices to show that $H(z_i)$ is a **P**-matrix for all $z_i \in \widetilde{\mathcal{Z}}$. Given $z_i \in \widetilde{\mathcal{Z}}$, let $H = H(z_i)$. Then,

$$H = H(z_i) = \begin{pmatrix} A_i & B \\ C_i & D \end{pmatrix}, \tag{3.36}$$

where $A_i = \sigma b^*(X_i)^{\sigma-2}\left[X_i(I + ee^T) + (\sigma-1)x_i e^T\right] + E_i$, $B = -\frac{1}{k+1}e$, $C_i = -\sigma b^*(X_i)^{\sigma-1}e^T$, and $D = \frac{1}{k+1}$, where $X_i = \sum_{j=1}^{N} x_{ij}$, $x_i = (x_{11}, \ldots, x_{1N})^T$, and E_i is an $N \times N$ diagonal matrix with $c_j''(x_{ij})$ as its jth diagonal entry. It suffices to show that H is a **P**-matrix when $E_i = 0$.

If $N < \frac{3\sigma-1}{\sigma-1}$, then A_i is positive semidefinite by Lemma 11. Therefore, we only consider the principal submatrix H_α of H, where $\alpha \subseteq \{1, \ldots, N\}$ is a nonempty index set, and $H_\alpha \triangleq \begin{pmatrix} A_\alpha & B_\alpha \\ C_\alpha & D \end{pmatrix}$, where $A_\alpha = \sigma b^*(X_i)^{\sigma-1}\left[I_{n_\alpha} + e^{n_\alpha}(e^{n_\alpha})^T\right] + \sigma(\sigma-1)b^*(X_i)^{\sigma-2}x_\alpha(e^{n_\alpha})^T, B_\alpha = -\frac{1}{k+1}e^{n_\alpha}, C_\alpha = -\sigma b^*(X_i)^{\sigma-1}(e^{n_\alpha})^T$, and I_{n_α} and e^{n_α} denote the identity matrix and the column of ones in $\mathbb{R}^{n_\alpha \times n_\alpha}$ and $\mathbb{R}^{n_\alpha}$, respectively, with $n_\alpha = |\alpha|$. Since

$$B_\alpha D^{-1} C_\alpha = \frac{1}{k+1}e^{n_\alpha}(k+1)\sigma b^*(X_i)^{\sigma-1}(e^{n_\alpha})^T = \sigma b^*(X_i)^{\sigma-1}e^{n_\alpha}(e^{n_\alpha})^T,$$

it follows that $A_\alpha - B_\alpha D^{-1} C_\alpha = \sigma b^*(X_i)^{\sigma-1} I_{n_\alpha} + \sigma(\sigma-1) b^*(X_i)^{\sigma-2} x_\alpha (e^{n_\alpha})^T$, which is a sum of a diagonal positive definite matrix and a $\mathbf{P_0}$-matrix, and thus is a $\mathbf{P}$-matrix. Therefore, $\det(H_\alpha) = \det(D)\det(A_\alpha - B_\alpha D^{-1} C_\alpha) > 0$ for all $\alpha \subseteq \{1, \ldots, N\}$ with $\alpha \neq \emptyset$, which implies that H is a $\mathbf{P}$-matrix. ∎

By leveraging Propositions 9 and 10, the convergence of the iterative fixed-point scheme can be claimed under the caveat that the aggregate output is always bounded away from zero, as stated by the next result, whose proof is similar to Theorem 10 and is omitted.

Corollary 4. *Suppose (A15), (A16) and (A17) hold. Suppose $X \geq \eta$ for some $\eta > 0$ and all $x \in K$, where $X = \sum_{i=1}^N x_i$. Let $\{x_i^k, \hat{\theta}_i^k\}$ be computed via Algorithm 5 for $i = 1, \ldots, N$. Suppose a unique solution to the fixed-point problem* (3.25) *can be obtained, given $p^k(\xi^k)$ and $\{\bar{\theta}_i^k\}_{i=1}^N$ for each $k \geq 0$. Then, $\hat{\theta}_i^k \to \theta^*$ almost surely for $i = 1, \ldots, N$ and $x_i^k \to x^*$ almost surely for $i = 1, \ldots, N$, where x^* is a solution of the variational inequality* (3.2).

We conclude this section with an observation. If one used a more widely used estimation technique such as a least-squares estimation then it remains unclear if almost-sure convergence statements can always be claimed since least-squares estimators generally converge in a weaker-sense while stronger statements may be available for linear regression (see [90]). In effect, a scheme that combines a least-squares estimation technique with a strategy update, while convergent, *may* not possess desirable almost-sure convergence properties. While, we examine nonlinear Nash-Cournot games in this section, we also show that such claims hold for more general aggregative Nash games. However, it should be emphasized that extending this avenue to Nash games where the associated variational map is non-monotone may lead to challenges. In particular, what are perfectly reasonable schemes for a subclass of Nash games may not be supported by similar asymptotic guarantees when the structural properties of the problem do not satisfy some key requirements.

3.4 Numerical results

In this section, we apply the developed algorithms on a class of networked Nash-Cournot games described in Section 3.4.1. In Section 3.4.2, we apply the distributed gradient-based schemes for purposes of learning equilibria and the misspecified parameters when aggregate output is observable, while in Section 3.4.2, we apply the proposed iterative fixed-point schemes when aggregate output is unobservable. Note that the simulations were carried out on Matlab R2009a on a laptop with Intel Core 2 Duo CPU (2.40GHz) and 2GB memory. The complementarity solver `PATH`, developed by [53], was utilized for solving the variational inequality problems that arose in implementing the algorithms.

3.4.1 Problem description

We consider a setting where there are N firms competing over a W-node network. Firm f may produce and sell its good at node i (denoted by g_{fi} and s_{fi}, respectively), where $f = 1, \ldots, N$ and $i = 1, \ldots, W$. We assume that for a given firm f, the cost of generating g_{fi} units of power at node i is linear and is given by $c_{fi}g_{fi}$. Furthermore, the generation level associated with firm f is bounded by its production capacity, which is denoted by cap_{fi}. The aggregate sales of all firms at node i is denoted by S_i, and the nodal price of power at node i, assumed to be a linear function of S_i, is defined as $p_i(S_i) \triangleq a_i^* - b_i^* S_i$, where a_i^* and b_i^* are node-specific positive price function parameters. A given firm can produce at any node and then sell at different nodes, provided that the aggregate production at all nodes matches the aggregate sales at all nodes for each firm. For simplicity, we assume that there is no transportation cost between any two nodes, and that there is no limit of sales at any node. Then, the resulting problem faced by firm f can be stated as

$$(\text{Firm}(x_{-f})) \quad \max_{s_{fi} \geq 0,\, \text{cap}_{fi} \geq g_{fi} \geq 0} \left\{ \sum_{i=1}^{W} (p_i(S_i) s_{fi} - c_{fi} g_{fi}) : \sum_{i=1}^{W} (s_{fi} - g_{fi}) = 0 \right\}. \tag{3.37}$$

The resulting Nash-Cournot equilibrium is given by $\{x_f^*\}_{f=1}^N$ where x_f^* is a solution to $(\text{Firm}(x_{-f}^*))$ for $f = 1, \ldots, N$. Prices are assumed to be corrupted by noise, in one of two ways:

$$p_i(S_i; \xi_i) = (a_i^* + \xi_i) - b_i^* S_i, \tag{3.38}$$

$$p_i(S_i; \xi_i) = a_i^* - (b_i^* + \xi_i) S_i. \tag{3.39}$$

Note that firm f either has to learn $\theta^* \triangleq (a_i^*)_{i=1}^W$ when prices are given by (3.38) or learn $\theta^* \triangleq (b_i^*)_{i=1}^W$ when prices are given by (3.39). In the remainder of this section, let $a^* \triangleq (a_1^*, \ldots, a_W^*)^T$, $b^* \triangleq (b_1^*, \ldots, b_W^*)^T$, $\theta^* \triangleq (\theta_1^*, \ldots, \theta_W^*)^T$, $\xi^* \triangleq (\xi_1^*, \ldots, \xi_W^*)^T$ and $x \triangleq (x_1^T, \ldots, x_W^T)^T$ with $x_i \triangleq (s_{1i}, s_{2i}, \ldots, s_{Ni}, g_{1i}, g_{2i}, \ldots, g_{Ni})^T$. Note that this problem is employed as a motivating example since Cournot-based models have been used extensively in their analysis (cf. [73, 74]). Naturally, a range of rationality assumptions can be imposed on firms in power markets, but given the sheer size of the problem and the repeated nature of competition (in most power markets, firms compete as many as 5–6 times every hour in the setting of prices) with relatively minor changes occuring in demand/availability over a short period.

3.4.2 Learning with observation of the aggregate output

In this subsection, we assume that every firm knows the aggregate output at each node, and employ the learning schemes proposed in Section 3.2.1. Suppose, the nodal price function is given by (3.38) and suppose Algorithm 4 (the gradient-based distributed learning scheme), proposed in Section 3.2.1, is employed for learning parameters and computing equilibria. Suppose firms have generated a price at each node. We use $p_i = a_i^* + \xi_i - b_i^* S_i$ to denote the price. Each firm will solve the following (regularized) problem to estimate a_i^* and b_i^*:

$$\min_{\{a_i, b_i\} \in \Theta} \quad \mathbb{E}\left[(a_i - b_i S_i - p_i)^2 + \lambda a_i^2 + \lambda b_i^2\right]. \tag{3.40}$$

Suppose S_i is as per a uniform distribution and is specified by $S_i \sim U[0, a_i^0/b_i^0]$, where a_i^0 and b_i^0 are initial estimates of a_i^* and b_i^*. Suppose, the noise ξ_i is distributed as per a uniform distribution and is specified by $\xi_i \sim U[-a^*/2, a^*/2]$. Suppose the steplength sequence $\{\gamma_i^k\}$ and $\{\alpha_i^k\}$ are chosen according to Lemma 9: $\gamma_i^k = \frac{1}{(k+N_i)^\alpha}$ and $\alpha_i^k = \frac{1}{(k+M_i)^\beta}$, where $\alpha = 0.8$ and $\beta = 0.6$ and N_i and M_i are randomly chosen from an interval $[1, 200]$. The algorithm was terminated at $k = 10000$. Table 3.1 shows the scaled errors of the learning scheme.

Table 3.1: Distributed gradient scheme

N	W	Learning a^* and b^*		
		$\frac{\|x^k - x^*\|}{1+\|x^*\|}$	$\frac{\|\hat{a}^k - a^*\|}{1+\|a^*\|}$	$\frac{\|\hat{b}^k - b^*\|}{1+\|b^*\|}$
5	1	7.2×10^{-7}	2.9×10^{-2}	4.7×10^{-2}
5	2	3.3×10^{-4}	3.3×10^{-2}	5.3×10^{-2}
5	3	7.4×10^{-5}	3.3×10^{-2}	5.3×10^{-2}
5	4	1.2×10^{-2}	4.2×10^{-2}	6.8×10^{-2}
5	5	1.4×10^{-2}	3.2×10^{-2}	8.5×10^{-2}
10	2	1.3×10^{-4}	3.4×10^{-2}	3.7×10^{-2}
10	4	1.1×10^{-2}	2.6×10^{-2}	8.4×10^{-2}
10	6	2.4×10^{-2}	3.6×10^{-2}	8.6×10^{-2}
10	8	2.8×10^{-2}	3.0×10^{-2}	6.4×10^{-2}
10	10	3.1×10^{-2}	4.1×10^{-2}	5.4×10^{-2}

Table 3.2: Iterative fixed-point scheme

N	W	Learning a^*		Learning b^*	
		$\max_f \frac{\|x_f^k - x^*\|}{1+\|x^*\|}$	$\max_f \frac{\|\hat{a}_f^k - a^*\|}{1+\|a^*\|}$	$\max_f \frac{\|x_f^k - x^*\|}{1+\|x^*\|}$	$\max_f \frac{\|\hat{b}_f^k - b^*\|}{1+\|b^*\|}$
5	1	6.0×10^{-3}	5.4×10^{-3}	2.3×10^{-3}	1.5×10^{-3}
5	2	1.9×10^{-3}	1.6×10^{-3}	9.1×10^{-4}	7.7×10^{-4}
5	3	1.4×10^{-3}	2.7×10^{-3}	7.8×10^{-4}	1.4×10^{-3}
5	4	7.8×10^{-3}	2.8×10^{-3}	2.0×10^{-3}	1.0×10^{-3}
5	5	1.0×10^{-3}	2.5×10^{-3}	1.2×10^{-2}	2.2×10^{-3}
10	2	2.0×10^{-3}	1.9×10^{-3}	1.2×10^{-3}	1.2×10^{-3}
10	4	1.1×10^{-2}	4.2×10^{-3}	1.5×10^{-2}	9.4×10^{-4}
10	6	1.8×10^{-3}	0.8×10^{-3}	3.0×10^{-4}	1.5×10^{-3}
10	8	2.0×10^{-3}	2.7×10^{-3}	1.3×10^{-3}	8.5×10^{-4}
10	10	1.1×10^{-3}	3.5×10^{-3}	3.8×10^{-4}	7.0×10^{-4}

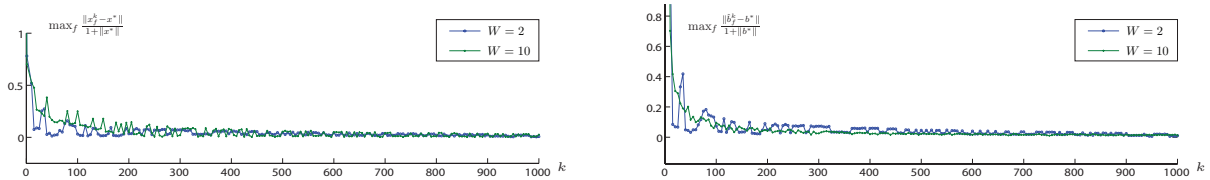
Learning without observing the aggregate output

In this subsection, we examine how the schemes perform when firms are ignorant of aggregate output at each node while a common knowledge assumption is assumed to hold.

Suppose, the nodal price function is given by (3.38) or (3.39) and suppose Algorithm 5 (the iterative fixed-point scheme), proposed in Section 3.3.1, is employed for learning parameters and computing equilibria. Suppose, the noise ξ is distributed as per a uniform distribution and is specified by $\xi \sim U[-\theta^*/2, \theta^*/2]$. Each run comprised of 10000 steps learning a^* and 50000 steps for learning b^*. Table 3.2 shows the scaled errors of the learning scheme while Figures 3.1(a) and 3.1(b) illustrate the scaled errors of the learning scheme when the number of steps, denoted by k, increases for learning x^* and a^*, respectively. Analogous figures for learning x^* and b^* are provided (see Figures 3.2(a) and 3.2(b)).

(a) Learning x (b) Learning a

Figure 3.1: Computing x^* and learning a^* ($\xi \sim U[-\theta^*/2, \theta^*/2]$, $N = 10$)



(a) Learning x (b) Learning b

Figure 3.2: Computing x^* and learning b^* ($\xi \sim U[-\theta^*/2, \theta^*/2]$, $N = 10$)

Table 3.3: Learning x^* and b^* in a stochastic regime when $N = 5$ and $W = 1$, stopping at $k = 10000$

(a) $\xi \sim U[-b^*/2, b^*/2]$

	Sequential		Simultaneous	
Bound	$\frac{\|x^k - x^*\|}{1+\|x^*\|}$	$\frac{\|\hat{b}^k - b^*\|}{1+\|b^*\|}$	$\frac{\max \|x_f^k - x^*\|}{1+\|x^*\|}$	$\frac{\max \|\hat{b}_f^k - b^*\|}{1+\|b^*\|}$
32.3664	2.1×10^{-1}	1.2×10^{-1}	4.9×10^{-3}	3.3×10^{-3}
64.7329	1.2×10^{-1}	1.0×10^{-1}	5.0×10^{-3}	3.3×10^{-3}
97.0993	5.5×10^{-1}	8.8×10^{-1}	5.0×10^{-3}	3.3×10^{-3}
129.4658	7.4×10^{-1}	1.1	5.1×10^{-3}	3.4×10^{-3}
161.8322	1.2	7.9×10^{-1}	5.1×10^{-3}	3.4×10^{-3}

(b) $\xi \sim U[-R, R]$

	Sequential		Simultaneous	
R	$\frac{\|x^k - x^*\|}{1+\|x^*\|}$	$\frac{\|\hat{b}^k - b^*\|}{1+\|b^*\|}$	$\frac{\max \|x_f^k - x^*\|}{1+\|x^*\|}$	$\frac{\max \|\hat{b}_f^k - b^*\|}{1+\|b^*\|}$
$b^*/5$	7.5×10^{-2}	4.8×10^{-2}	1.9×10^{-3}	1.2×10^{-3}
$b^*/4$	9.6×10^{-2}	6.0×10^{-2}	2.4×10^{-3}	1.6×10^{-3}
$b^*/3$	1.3×10^{-1}	8.0×10^{-2}	3.2×10^{-3}	2.2×10^{-3}
$b^*/2$	2.1×10^{-1}	1.2×10^{-1}	4.9×10^{-3}	3.3×10^{-3}
$b^*/1$	5.3×10^{-1}	2.3×10^{-1}	9.9×10^{-3}	6.7×10^{-3}

In Table 3.3(a), we raise the upper bounds of the strategy sets of all agents and compare a sequential scheme with our iterative fixed-point scheme. In the sequential counterpart, we employ $10,000$ steps of stochastic approximation-based learning followed by $10,000$ steps of computation. It is seen that the error from the sequential scheme increases proportionally to the bound, while the error associated with our simultaneous scheme does not change significantly. Table 3.3(b) shows that when increasing the variance of the noise makes the difference in errors between the sequential and simultaneous schemes more pronounced. Consequently, for the same effort, it can be seen that the simultaneous scheme performs far better to the

sequential scheme, particularly when the variance of the noise grows.

3.5 Concluding remarks

Nash games, a broadly applicable paradigm for modeling strategic interactions in noncooperative settings, have emerged as immensely useful in the context of distributed control problems. Yet, the development of distributed protocols for learning equilibria may be complicated by several challenges: (i) Agents may have an incomplete specification of payoffs; (ii) Agents may be unavailable to observe the actions of their counterparts; and finally, (iii) Observations may be corrupted by noise. Accordingly, this chapter is motivated by developing schemes for learning equilibria and resolving misspecification (such as in the price functions). We consider two specific settings as part of our investigation and apply these techniques on a class of networked Nash-Cournot games. First, we consider convex static stochastic Nash games characterized by a suitable monotonicity property in which agent payoffs are parameterized by a misspecified vector. We consider a framework that combines (stochastic) gradient steps with a stochastic approximation step that attempts to learn the parameter. In such settings, we provide asymptotic statements that show that agents may learn equilibria and the true parameters in an almost sure sense. In addition, we provide non-asymptotic error bounds that demonstrate that the rate of convergence is not impaired by the presence of learning. Second, we refine our statements to a Cournot regime where we assume *common knowledge* holds but aggregate output is unobservable. In such a setting, we construct a learning scheme in which firms maintain a belief of the aggregate output and the misspecified price function parameter. After each step, these beliefs are updated by employing fixed-point steps and by leveraging the disparity between estimated and (noisy) observed prices. We proceed to show that in the limit, every firm learns the true Nash-Cournot equilibrium strategy in an almost-sure sense. Additionally, every firm learns the correct value of the misspecified parameter in an almost-sure sense. Yet much remains to be studied, including weakening monotonicity requirements on the map and boundedness requirements on the strategy sets. It also remains to be investigated as to whether learning can allow for weakening the common knowledge assumption.

Chapter 4

Misspecified Markov Decision Processes

4.1 Introduction

Markov decision processes (MDPs) are an important class of models for analyzing dynamic decision making problems. First examined in [91], such models have been used in a number of domains including robotics, control-theory, economics, healthcare, and manufacturing. Specifically, a Markov decision process is a discrete time stochastic control process. At each time step, the process is in some state s, and the decision maker may choose an action a that is available in state s. The process responds at the next time step by moving to a new state s', and giving the decision maker a corresponding reward $R_a(s, s')$ or cost $C_a(s, s')$. The next state s' depends on the current state s and the decision maker's action a, but given s and a, it is conditionally independent of all previous states and actions; in other words, the state transitions of an MDP have the Markov property. In an MDP with a discrete state space, the state transition probabilities from time t to $t+1$ are specified by an action U_t at time t, i.e., $\mathbb{P}(s' \mid s, a) \triangleq \mathbb{P}(X_{t+1} = s' \mid X_t = s, U_t = a)$, where at time t, X_t and U_t denotes the state of the process and the transition matrix, respectively.

Suppose $\mathcal{A}$ and $\mathcal{S}$ denote the set of actions and states. Suppose $C(a, s; \psi^*)$ denotes the correctly specified cost of taking action a at state s where $\gamma \in [0, 1)$ denotes the discount factor. The probability of the system transitioning from state s' to s'' based on action a is specified by $\mathbb{P}^*(s = s'' \mid s = s', a)$. Furthermore, we define a policy map as $\pi : \mathcal{S} \to \mathcal{A}$ while the value function of a policy π is denoted by $V^\pi : \mathcal{S} \to \mathbb{R}$ and $V^\pi(s)$ denotes the expected discounted cost of policy π when starting at state s. The objective lies in determining a policy π that minimizes the discounted expected sum over an infinite horizon, given by $\sum_{k=0}^{\infty} \gamma^k C(s_k, a_k; \psi^*)$, where $a_{k+1} = \pi(s_k)$.

This chapter considers the resolution of such problems in regimes where the transition matrix $\mathbb{P}^*$ and the parametrization of the cost function ψ^* are unavailable a priori. Estimation of transition matrices has been studied extensively in the literature [92, 93, 94] while robust optimization approaches have also been employed (cf. [95, 96]). A rather distinct approach in contending with the absence of information is embodied by the Q-learning algorithm presented in [97]. This is a simulation-based technique for computing estimates

to the value function and has a similar structure to stochastic approximation algorithms [98]. Simulation-based approaches have also been reviewed in [99], particularly notable being the upper confidence bound (UCB) sampling algorithm (cf. [100, 101, 102]).

Given an MDP($\mathbb{P}^*, \psi^*$) where $\mathbb{P}^*$ and ψ^* are unavailable, a standard approach is the following:

$$\text{(1) Learn } \mathbb{P}^* \text{ and } \psi^*; \qquad \text{(2) Solve MDP}(\mathbb{P}^*, \psi^*).$$

This technique is afflicted by several challenges, a subset of which we describe next:

(i) *Inability to accommodate streaming data:* Increasingly, MDP-based models have to be constantly updated with new, and possibly, streaming data. Yet the traditionally developed asymptotics and error analysis for resolving MDPs cannot accommodate streaming data.

(ii) *Lack of asymptotics:* Step (1) often requires solving stochastic and/or large-scale learning problems whose solutions are obtained in an asymptotic sense. Any practical implementation of this scheme necessitates that Step (1) terminate finitely; however, premature termination of (1) leads to estimators afflicted by error and may result in significant error in the computed value function. In effect, asymptotic convergence of this scheme cannot be claimed.

(iii) *Practical implementations:* Step (1) may require infinite time, particularly since it requires solving stochastic optimization problems and during this period, no estimate of the optimal value function is available. In effect, error bounds can only be prescribed after step (1) is complete.

A simultaneous scheme for learning and computation: We consider an avenue that has found recent application for resolving misspecified optimization and variational problems in stochastic regimes [79, 103]. This necessitates a *simultaneous* approach in which the learning problems for $\mathbb{P}^*$ and ψ^* are resolved simultaneously with the original MDP. In effect, we consider the estimators from the coupled dynamics and examine both the asymptotics and error bounds for a variety of computational schemes. Our scheme relies on the prescription of learning problems.

(i) *Learning of transition matrices:* We consider the following learning problem for transition matrices based on using observational data:

$$\mathbb{P}^* \in \operatorname*{argmin}_{\mathbb{P}\in\mathcal{P}} \mathbb{E}[g(\mathbb{P};\eta)], \tag{$\mathcal{L}^{\mathbb{P}}$}$$

where $\mathcal{P}$ denotes the space of stochastic matrices, i.e. nonnegative matrices with row sums equal to unity.

(ii) *Misspecification of cost functions:* The cost functions are parameterized by a vector ψ^*, representing a set of parameters idiosyncratic to the machine of interest. For instance, it may pertain to the efficiency of the

machine, the start-up/shut-down times, the skill of the workers in question etc. All of these parameters may require learning, often via an online approach that incorporates the use of observations, possibly corrupted by noise. Such a problem can be cast as a stochastic optimization problem, defined as follows:

$$\psi^* \in \operatorname*{argmin}_{\psi \in \Psi} \mathbb{E}[R(\psi; \xi)], \tag{$\mathcal{R}^\Psi$}$$

where ξ a random variable and Ψ denotes the feasibility set for ψ. By using stochastic approximation, we may generate sequences $\{\mathbb{P}_k\}$ and $\{\psi_k\}$ such that $\mathbb{P}_k \to \mathbb{P}^*$ and $\psi_k \to \psi^*$ as $k \to \infty$ in an a.s. sense.

We provide an illustration of the approach by using the well-studied value iteration scheme as a basis [104]. In its original form, value iteration maintains an estimate of the value function and updates this belief based on solving a suitable problem. When the change in the value functions falls within a suitably defined threshold in a norm-sense, the scheme terminates. We now provide a relatively quick overview of this scheme (cf. [3]). Let $\mathcal{V}$ denote the space of value functions and $\mathcal{M} : \mathcal{V} \to \mathcal{V}$ be a mapping such that for each $s \in \mathcal{S}$, $\mathcal{M}$ is defined as follows:

$$\mathcal{M}v(s) = \max_{a \in \mathcal{A}} \left\{ C(s, a; \psi^*) + \gamma \sum_{s' \in \mathcal{S}} \mathbb{P}^*(s' \mid s, a) v(s') \right\}. \tag{MDP$(\mathbb{P}^*, \Psi^*)$}$$

Given a v^0, the value iteration scheme is defined as follows:

$$v^{k+1} := \mathcal{M}v^k, \qquad \text{for } k \geq 0. \tag{Value Iteration}$$

Since $\mathcal{M}$ is a contraction mapping on $\mathcal{V}$ if $0 \leq \gamma < 1$ (cf. Proposition 3.10.2 in [3]), convergence of the scheme can be shown within a reasonably straightforward fashion. However, one of the challenges lies in the availability of $C(s, a; \psi^*)$ and $\mathbb{P}^*$, motivating the development of a misspecified variant. We assume the cost and matrix to be given by $C(s, a; \widetilde{\psi})$ and $\widetilde{\mathbb{P}}(s'|s, a)$. We may then define a *misspecified* operator $\widetilde{\mathcal{M}}_k : \mathcal{V} \to \mathcal{V}$ by utilizing estimates $\widetilde{\mathbb{P}}_k$ and $\widetilde{\psi}_k$:

$$\widetilde{\mathcal{M}}_k v(s) = \max_{a \in \mathcal{A}} \left\{ C(s, a; \widetilde{\psi}_k) + \gamma \sum_{s' \in \mathcal{S}} \widetilde{\mathbb{P}}_k(s'|s, a) v(s') \right\}. \tag{4.1}$$

Specifically, we assume that $\widetilde{\mathbb{P}}_k$ and $\widetilde{\psi}_k$ are sequences converging to $\mathbb{P}^*$ and ψ^* *a.s.* as a result of stochastic approximation schemes.

We now present our main research questions and provide an outline of this chapter:

(i) Misspecified value iteration: In Section 2, we present a misspecified value iteration scheme for addressing MDPs in which the cost function and transition matrices are misspecified. We examine the asymptotics

of the resulting scheme and providing a quantification of the degradation of the rate of convergence based on the presence of learning.

(ii) Misspecified policy iteration: In Section 3, we consider an analogous set of questions in the regime of policy iteration where we provide almost sure convergence statements.

(iii) Misspecified Q-learning: Finally, in Section 4, we consider Q-learning approaches for solving MDPs with misspecified cost functions and present constant steplength error bounds for extensions that resolve the misspecification while solving the original MDP.

4.2 Misspecified value iteration

Value iteration [91] represents amongst the oldest of schemes for solving an MDP. We begin by presenting a *misspecified* value iteration scheme for resolving MDP$(\mathbb{P}^*, \psi^*)$ and subsequently present asymptotic convergence and error analysis.

We define $\mathcal{P}$ to be set of all transition matrices, vec$(\mathbb{P})$ to be the vector drawn from the entries of $\mathbb{P}$ for all $\mathbb{P} \in \mathcal{P}$, and vec$(\mathcal{P}) \triangleq \{\text{vec}(\mathbb{P}) : \mathbb{P} \in \mathcal{P}\}$. Estimating $\mathbb{P}^*$ often requires the resolution of a suitably defined learning problem, given by a stochastic optimization problem $(\mathcal{L}^{\mathbb{P}})$, where vec$(\mathcal{P})$ is a closed and convex set, $\eta : \Lambda \to \mathbb{R}^p$ is a random variable defined on a probability space $(\Lambda, \mathcal{F}_\eta, \mathbb{P}_\eta)$, and $g : \mathbf{P} \times \Lambda \to \mathbb{R}$ is a real-valued function. We may specify our joint scheme for learning and computation as follows:

Algorithm 6 (**Misspecified Value Iteration**). **Step 0:** *Let* $\tilde{v}^0 : \mathcal{S} \to \mathbb{R}$, $vec(\widetilde{\mathbb{P}}_0) \in vec(\mathbf{P})$, $\widetilde{\psi}_0 \in \Psi$, $\alpha_0 > 0$, $\beta_0 > 0$ *and* $k = 0$.

Step 1: *For all* $k \geq 0$,

$$\tilde{v}^{k+1} := \widetilde{\mathcal{M}}_k \tilde{v}^k, \qquad \text{(Computation)}$$

$$vec(\widetilde{\mathbb{P}}_{k+1}) := \Pi_{vec(\mathbf{P})} \left(vec(\widetilde{\mathbb{P}}_k) - \alpha_k (\nabla g(\widetilde{\mathbb{P}}_k) + w_k) \right), \qquad \text{(Learning}-\mathbb{P}\text{)}$$

$$\widetilde{\psi}_{k+1} := \Pi_{\Psi} \left(\widetilde{\psi}_k - \beta_k (\nabla R(\widetilde{\psi}_k) + u_k) \right), \qquad \text{(Learning}-\psi\text{)}$$

where $w_k \triangleq \nabla g(\widetilde{\mathbb{P}}_k; \eta_k) - \nabla g(\widetilde{\mathbb{P}}_k)$, $g(\mathbb{P}) \triangleq \mathbb{E}[g(\mathbb{P}; \eta)]$, $u_k = \nabla R(\widetilde{\psi}_k; \xi_k) - \nabla R(\widetilde{\psi}_k)$, $R(\psi) \triangleq \mathbb{E}[R(\psi; \xi)]$, $\widetilde{\mathcal{M}}_k v(s) := \max_{a \in \mathcal{A}} (C(s, a; \widetilde{\psi}_k) + \gamma \sum_{s' \in \mathcal{S}} \widetilde{\mathbb{P}}_n(s'|s, a) v(s'))$, *and* α_k *and* β_k *are chosen according to Proposition 11.*

Step 2: *If* $k > K$, *stop; else* $k := k + 1$ *and go to Step 1.*

We begin by showing that the misspecified operator $\widetilde{\mathcal{M}_k}$ is a contraction mapping for any k. We suppress the subscript k in this proof for purposes of clarity.

Lemma 12 (Contractive property of $\widetilde{\mathcal{M}}$)**.** *Define $\widetilde{\mathcal{M}}$ by suppressing the subscript k in* (4.1)*, i.e.*

$$\widetilde{\mathcal{M}}v(s) = \max_{a\in\mathcal{A}} \left\{ C(s,a;\widetilde{\psi}) + \gamma \sum_{s'\in\mathcal{S}} \widetilde{\mathbb{P}}(s'|s,a)v(s') \right\}.$$

If $0 \leq \gamma < 1$, then $\widetilde{\mathcal{M}}$ is a contraction mapping on $\mathcal{V}$.

Proof. Let $u, v \in \mathcal{V}$ and assume that $\widetilde{\mathcal{M}}v(s) \geq \widetilde{\mathcal{M}}u(s)$ without loss of generality for any state s. For any state s, let $\tilde{a}_s^*(v)$ be defined as follows:

$$\tilde{a}_s^*(v) = \operatorname*{argmax}_{a\in\mathcal{A}} \left\{ C(s,a;\widetilde{\psi}) + \gamma \sum_{s'\in\mathcal{S}} \widetilde{\mathbb{P}}(s'|s,a)v(s') \right\}.$$

Then, we have the following sequence of inequalities:

$$\begin{aligned} 0 \leq \widetilde{\mathcal{M}}v(s) - \widetilde{\mathcal{M}}u(s) &= C(s,\tilde{a}_s^*(v);\widetilde{\psi}) + \gamma \sum_{s'\in\mathcal{S}} \widetilde{\mathbb{P}}(s'|s,\tilde{a}_s^*(v))v(s') - \left(C(s,\tilde{a}_s^*(u);\widetilde{\psi}) + \gamma \sum_{s'\in\mathcal{S}} \widetilde{\mathbb{P}}(s'|s,\tilde{a}_s^*(u))u(s') \right) \\ &\leq \underbrace{C(s,\tilde{a}_s^*(v);\widetilde{\psi}) + \gamma \sum_{s'\in\mathcal{S}} \widetilde{\mathbb{P}}(s'|s,\tilde{a}_s^*(v))v(s') - \left(C(s,\tilde{a}_s^*(v);\widetilde{\psi}) + \gamma \sum_{s'\in\mathcal{S}} \widetilde{\mathbb{P}}(s'|s,\tilde{a}_s^*(v))u(s') \right)}_{\text{Term (a)}}, \end{aligned}$$

where the second inequality is a consequence of noting that for all s, we have the following:

$$\begin{aligned} \widetilde{\mathcal{M}}u(s) = \max_{a\in\mathcal{A}} \left\{ C(s,a;\widetilde{\psi}) + \gamma \sum_{s'\in\mathcal{S}} \widetilde{\mathbb{P}}(s'|s,a)u(s') \right\} &= \left(C(s,\tilde{a}_s^*(u);\widetilde{\psi}) + \gamma \sum_{s'\in\mathcal{S}} \widetilde{\mathbb{P}}(s'|s,\tilde{a}_s^*(u))u(s') \right) \\ &\geq \left(C(s,\tilde{a}_s^*(v);\widetilde{\psi}) + \gamma \sum_{s'\in\mathcal{S}} \widetilde{\mathbb{P}}(s'|s,\tilde{a}_s^*(v))u(s') \right). \end{aligned}$$

It follows that Term (a) can be bounded as follows:

$$\begin{aligned} &C(s,\tilde{a}_s^*(v);\widetilde{\psi}) + \gamma \sum_{s'\in\mathcal{S}} \widetilde{\mathbb{P}}(s'|s,\tilde{a}_s^*(v))v(s') - \left(C(s,\tilde{a}_s^*(v);\widetilde{\psi}) + \gamma \sum_{s'\in\mathcal{S}} \widetilde{\mathbb{P}}(s'|s,\tilde{a}_s^*(v))u(s') \right) \\ &= \gamma \sum_{s'\in\mathcal{S}} \widetilde{\mathbb{P}}(s'|s,\tilde{a}_s^*(v)) \left(v(s') - u(s')\right) \leq \gamma \sum_{s'\in\mathcal{S}} \widetilde{\mathbb{P}}(s'|s,\tilde{a}_s^*(v)) \|v-u\|_\infty = \gamma \|v-u\|_\infty, \end{aligned}$$

Consequently, $\|\widetilde{\mathcal{M}}v - \widetilde{\mathcal{M}}u\|_\infty = \sup_{s\in\mathcal{S}} |\widetilde{\mathcal{M}}v(s) - \widetilde{\mathcal{M}}u(s)| \leq \gamma \|v-u\|_\infty$, implying that $\widetilde{\mathcal{M}}$ is contractive. ∎

Our next proposition shows that when the estimated transition matrix is within some bound of its true counterpart, under a suitable Lipschitzian requirement of $C(s,a,\psi)$ in ψ, we obtain the following relationship between the true operator and its misspecified counterpart. This lemma subsequently finds application in the main convergence result.

Lemma 13. *Suppose* $\sum_{s' \in \mathcal{S}} |\mathbb{P}^*(s'|s,a) - \widetilde{\mathbb{P}}(s'|s,a)| \leq \delta$ *for all* s *and* a. *Suppose* $C(s,a;\psi)$ *is Lipschitz continuous in* ψ *with constant* L_C *uniformly in* s *and* a. *Then the following holds for all* $u, v \in \mathcal{V}$*:*

$$\|\mathcal{M}v - \widetilde{\mathcal{M}}u\| \leq L_C\|\psi^* - \widetilde{\psi}\| + \gamma\delta(\|v\| + \|u\|) + \gamma\|v - u\|.$$

Proof. Let $u, v \in \mathcal{V}$ and assume without loss of generality that $\mathcal{M}v(s) \geq \widetilde{\mathcal{M}}u(s)$. For a state s, we may define $a_s^*(v)$ and $\tilde{a}_s^*(v)$ as follows:

$$a_s^*(v) \triangleq \operatorname*{argmax}_{a \in \mathcal{A}}(C(s,a;\psi^*) + \gamma \sum_{s' \in \mathcal{S}} \mathbb{P}^*(s'|s,a)v(s')) \text{ and } \tilde{a}_s^*(v) \triangleq \operatorname*{argmax}_{a \in \mathcal{A}}(C(s,a;\widetilde{\psi}) + \gamma \sum_{s' \in \mathcal{S}} \widetilde{\mathbb{P}}(s'|s,a)v(s')).$$

Then, we have the following set of relations:

$$\begin{aligned}
0 &\leq \mathcal{M}v(s) - \widetilde{\mathcal{M}}u(s) = \mathcal{M}v(s) - \widetilde{\mathcal{M}}v(s) + \widetilde{\mathcal{M}}v(s) - \widetilde{\mathcal{M}}u(s) \\
&= C(s, a_s^*(v); \psi^*) + \gamma \sum_{s' \in \mathcal{S}} \mathbb{P}^*(s'|s, a_s^*(v))v(s') - \left(C(s, \tilde{a}_s^*(v); \widetilde{\psi}) + \gamma \sum_{s' \in \mathcal{S}} \widetilde{\mathbb{P}}(s'|s, \tilde{a}_s^*(v))v(s') \right) \\
&\quad + \widetilde{\mathcal{M}}v(s) - \widetilde{\mathcal{M}}u(s) \\
&\leq C(s, a_s^*(v); \psi^*) + \gamma \sum_{s' \in \mathcal{S}} \mathbb{P}^*(s'|s, a_s^*(v))v(s') - \left(C(s, a_s^*(v); \widetilde{\psi}) + \gamma \sum_{s' \in \mathcal{S}} \widetilde{\mathbb{P}}(s'|s, a_s^*(v))v(s') \right) \\
&\quad + \widetilde{\mathcal{M}}v(s) - \widetilde{\mathcal{M}}u(s),
\end{aligned}$$

where the second inequality follows from the suboptimality of $a_s^*(v)$ with respect to $a^*(v)$. It follows that

$$\begin{aligned}
&C(s, a_s^*(v); \psi^*) + \gamma \sum_{s' \in \mathcal{S}} \mathbb{P}^*(s'|s, a_s^*(v))v(s') - \left(C(s, a_s^*(v); \widetilde{\psi}) + \gamma \sum_{s' \in \mathcal{S}} \widetilde{\mathbb{P}}(s'|s, a_s^*(v))v(s') \right) \\
&+ \widetilde{\mathcal{M}}v(s) - \widetilde{\mathcal{M}}u(s) \\
&\leq L_C\|\psi^* - \widetilde{\psi}\| + \gamma \sum_{s' \in \mathcal{S}} \left(\mathbb{P}^*(s'|s, a_s^*(v)) - \widetilde{\mathbb{P}}(s'|s, a_s^*(v)) \right) v(s') + \widetilde{\mathcal{M}}v(s) - \widetilde{\mathcal{M}}u(s) \\
&\leq L_C\|\psi^* - \widetilde{\psi}\| + \gamma \sum_{s' \in \mathcal{S}} |\mathbb{P}^*(s'|s, a_s^*(v)) - \widetilde{\mathbb{P}}(s'|s, a_s^*(v))| \|v\| + \widetilde{\mathcal{M}}v(s) - \widetilde{\mathcal{M}}u(s) \\
&\leq L_C\|\psi^* - \widetilde{\psi}\| + \gamma\delta\|v\| + \widetilde{\mathcal{M}}v(s) - \widetilde{\mathcal{M}}u(s) \\
&\leq L_C\|\psi^* - \widetilde{\psi}\| + \gamma\delta\|v\| + \gamma\|v - u\| \\
&\leq L_C\|\psi^* - \widetilde{\psi}\| + \gamma\delta(\|v\| + \|u\|) + \gamma\|v - u\|,
\end{aligned}$$

where the first inequality follows from the Lipschitz continuity of $C(s,a;\psi)$ in ψ for fixed s and a, the second inequality is a consequence of the Cauchy-Schwarz inequality and the last inequality is a consequence of

invoking the contractive property of $\widetilde{\mathcal{M}}$ with constant γ. ∎

We are now ready to prove our main convergence statement.

Proposition 11 (**Misspecified value iteration: a.s. convergence and rate statement**). *Suppose $\{\tilde{v}^k\}$, $\{\widetilde{\mathbb{P}}_k\}$ and $\{\widetilde{\psi}_k\}$ are generated from Algorithm 6. Suppose the learning function $g(\cdot)$ is strongly convex in vec$(\mathbf{P})$, and the learning function $R(\cdot)$ is strongly convex in Ψ. Suppose $\alpha_k = \theta_1/k$ and $\beta_k = \theta_2/k$ with $\theta_1 > 1/(2\mu_g)$, $\theta_2 > 1/(2\mu_R)$, μ_g is the strong convexity constant of g and μ_R is the strong convexity constant of R. Suppose $C(s,a;\psi)$ is Lipschitz continuous in ψ with constant L_C for all s and a. Then, there exists a constant λ such that the following hold:*

(i) $\|\tilde{v}^k - v^*\| \to 0$, $\widetilde{\mathbb{P}}_k \to \mathbb{P}^*$ *and* $\widetilde{\psi}_k \to \psi^*$ *a.s. as* $k \to \infty$.

(ii) For any k, we have that the following holds:

$$\mathbb{E}\left[\|\tilde{v}^{k+1} - v^*\|\right] \leq \gamma^k \mathbb{E}[\|\tilde{v}^0 - v^*\|] + \sum_{j=1}^{k} \frac{\gamma^{k-j}\lambda}{\sqrt{j}} = \mathcal{O}\left(\frac{1}{\sqrt{k}}\right).$$

Proof. **(i)** First, we have that the following holds almost surely:

$$\begin{aligned}
\|\tilde{v}^{k+1} - v^*\| &= \|\widetilde{\mathcal{M}}_k \tilde{v}^k - \mathcal{M} v^*\| \\
&= \|\widetilde{\mathcal{M}}_k \tilde{v}^k - \widetilde{\mathcal{M}}_k v^* + \widetilde{\mathcal{M}}_k v^* - \mathcal{M} v^*\| \leq \|\widetilde{\mathcal{M}}_k \tilde{v}^k - \widetilde{\mathcal{M}}_k v^*\| + \|\widetilde{\mathcal{M}}_k v^* - \mathcal{M} v^*\| \\
&\leq \gamma \|\tilde{v}^k - v^*\| + L_C \|\widetilde{\psi}_k - \psi^*\| + \gamma \|\mathrm{vec}(\widetilde{\mathbb{P}}_k) - \mathrm{vec}(\mathbb{P}^*)\| \|v^*\|, \qquad (4.2)
\end{aligned}$$

where the last inequality follows from invoking Lemmas 12 and 13. Let $a_k = L_C \|\widetilde{\psi}_k - \psi^*\| + \gamma \|\mathrm{vec}(\widetilde{\mathbb{P}}_k) - \mathrm{vec}(\mathbb{P}^*)\| \|v^*\|$. Then, we have

$$\begin{aligned}
\|\tilde{v}^{k+1} - v^*\| &\leq \gamma \|\tilde{v}^k - v^*\| + a_k \leq \gamma(\gamma \|\tilde{v}^{k-1} - v^*\| + a_{k-1}) + a_k \\
&= \gamma^2 \|\tilde{v}^{k-1} - v^*\| + \gamma a_{k-1} + a_k \leq \cdots \leq \gamma^{k+1} \|\tilde{v}^0 - v^*\| + \sum_{i=0}^{k} \gamma^i a_{k-i}.
\end{aligned}$$

Since $\gamma^{k+1} \to 0$, it suffices to show that $\sum_{i=0}^{k} \gamma^i a_{k-i} \to 0$ as $k \to \infty$ in an a.s. sense. Since the learning problems for ψ^* and $\mathbb{P}^*$ are both strongly convex, we have that $a_k \to 0$ *a.s.* as $k \to \infty$. Then, for almost all $\omega \in \Omega$, given $\epsilon > 0$, there exists $N_1(\omega)$ such that $a_k \leq \epsilon$ for all $k \geq N_1(\omega)$. Also, for almost every $\omega \in \Omega$,

$a_k \leq L(\omega)$ for all k and some constant $L(\omega) > 0$. Thus, for $k \geq N_1(\omega)$,

$$\begin{aligned}\sum_{i=0}^{k} \gamma^i a_{k-i} &= \gamma^n a_0 + \ldots + \gamma^{k-N_1(\omega)} a_{N_1(\omega)} + \gamma^{k-N_1(\omega)-1} a_{N_1(\omega)+1} + \ldots + \gamma^0 a_k \\ &\leq (\gamma^k + \ldots + \gamma^{k-N_1(\omega)}) L(\omega) + \frac{\epsilon}{1-\gamma}.\end{aligned}$$

Since $\gamma^k \to 0$, there exists $N_2(\omega)$ such that $\gamma^k \leq \frac{\epsilon}{N_1(\omega)+1}, \ldots, \gamma^{k-N_1(\omega)} \leq \frac{\epsilon}{N_1(\omega)+1}$ for all $k \geq N_2(\omega)$. So, when $k \geq N(\omega) \triangleq \max\{N_1(\omega), N_2(\omega)\}$, we have that

$$\sum_{i=0}^{k} \gamma^i a_{k-i} \leq L(\omega)\epsilon + \frac{\epsilon}{1-\gamma} = \left(L(\omega) + \frac{1}{1-\gamma}\right)\epsilon.$$

Since $L(\omega)$ is finite in an a.s. sense and ϵ is arbitrarily chosen, proving that $\sum_{i=0}^{k} \gamma^i a_{n-i} \to 0$ *a.s.*. We may then conclude that $\|\tilde{v}^{k+1} - v^*\| \to 0$ in an a.s. sense as $k \to \infty$.

(ii) By taking expectations on both sides of (4.2), we have the following:

$$\mathbb{E}[\|\tilde{v}^{k+1} - v^*\|] \leq \gamma \mathbb{E}[\|\tilde{v}^k - v^*\|] + L_C \mathbb{E}[\|\widetilde{\psi}_k - \psi^*\|] + \gamma \mathbb{E}[\|\mathrm{vec}(\widetilde{\mathbb{P}}_k) - \mathrm{vec}(\mathbb{P}^*)\|]\|v^*\|. \tag{4.3}$$

Recall that the learning problem for ψ^* and $\mathbb{P}^*$ are both strongly convex. Then, we may use the standard rate estimate (see (5.292) in [42]) to get the following for suitably chosen λ_1 and λ_2:

$$\mathbb{E}[\|\widetilde{\psi}_k - \psi^*\|] \leq \frac{\lambda_1}{\sqrt{k}} \text{ and } \mathbb{E}[\|\mathrm{vec}(\widetilde{\mathbb{P}}_k) - \mathrm{vec}(\mathbb{P}^*)\|] \leq \frac{\lambda_2}{\sqrt{k}}. \tag{4.4}$$

Consequently, we obtain the following:

$$\mathbb{E}[\|\tilde{v}^{k+1} - v^*\|] \leq \gamma \mathbb{E}[\|\tilde{v}^k - v^*\|] + \frac{L_C\lambda_1 + \gamma\lambda_2\|v^*\|}{\sqrt{k}}.$$

Let $\lambda = L_C\lambda_1 + \gamma\lambda_2\|v^*\|$. Then, we have

$$\begin{aligned}\mathbb{E}[\|\tilde{v}^{k+1} - v^*\|] &\leq \gamma \mathbb{E}[\|\tilde{v}^k - v^*\|] + \frac{\lambda}{\sqrt{k}} \leq \gamma^2 \mathbb{E}[\|\tilde{v}^{k-1} - v^*\|] + \frac{\gamma\lambda}{\sqrt{k-1}} + \frac{\lambda}{\sqrt{k}} \\ &\leq \gamma^k \mathbb{E}[\|\tilde{v}^0 - v^*\|] + \sum_{j=1}^{k} \frac{\gamma^{k-j}\lambda}{\sqrt{j}}.\end{aligned}$$

Since γ^{-j} is increasing in j and $\frac{1}{\sqrt{j}}$ is decreasing in j, then there exists a K_1 such that $\frac{\gamma^{-j}}{\sqrt{j}}$ is decreasing in j for $j \leq K_1$ and $\frac{\gamma^{-j}}{\sqrt{j}}$ is increasing in j for $j > K_1$. Then, the second term in the above inequality can be

bounded as

$$\begin{aligned}\sum_{j=1}^{k}\frac{\gamma^{k-j}\lambda}{\sqrt{j}} &= \gamma^k\lambda\sum_{j=1}^{k}\frac{\gamma^{-j}}{\sqrt{j}} = \gamma^k\lambda\left(\sum_{j=1}^{K_1}\frac{\gamma^{-j}}{\sqrt{j}} + \sum_{j=K_1+1}^{k}\frac{\gamma^{-j}}{\sqrt{j}}\right)\\ &\le \gamma^k\lambda\left(K_1\gamma^{-1} + \int_{K_1+2}^{k+1}\frac{\gamma^{-t}}{\sqrt{t}}dt\right).\end{aligned} \tag{4.5}$$

Note that there exists a $K_2 > K_1+2$ such that $\frac{\gamma^{-t}}{t\sqrt{t}}$ is decreasing for $t \le K_2$ and $\frac{\gamma^{-t}}{t\sqrt{t}}$ is increasing for $t > K_2$. Then, we have

$$\begin{aligned}\int_{K_1+2}^{k+1}\frac{\gamma^{-t}}{\sqrt{t}}dt &= \frac{\gamma^{-t}}{\ln\gamma^{-1}}\cdot\frac{1}{\sqrt{t}}\Big|_{K_1+2}^{k+1} + \frac{1}{2}\int_{K_1+2}^{k+1}\frac{\gamma^{-t}}{\ln\gamma^{-1}}\cdot\frac{1}{t\sqrt{t}}dt\\ &\le \frac{1}{\ln\gamma^{-1}}\cdot\frac{\gamma^{-(k+1)}}{\sqrt{k+1}} + \frac{1}{2}\int_{K_1+2}^{K_2}\frac{\gamma^{-t}}{\ln\gamma^{-1}}\cdot\frac{1}{t\sqrt{t}}dt + \frac{1}{2}\int_{K_2}^{k+1}\frac{\gamma^{-t}}{\ln\gamma^{-1}}\cdot\frac{1}{t\sqrt{t}}dt\\ &\le \frac{1}{\ln\gamma^{-1}}\cdot\frac{\gamma^{-(k+1)}}{\sqrt{k+1}} + \frac{1}{2}\int_{K_1+2}^{K_2}\frac{\gamma^{-t}}{\ln\gamma^{-1}}\cdot\frac{1}{t\sqrt{t}}dt + \frac{\gamma^{-(k+1)}}{2\ln\gamma^{-1}}\cdot\frac{1}{(k+1)\sqrt{k+1}}\cdot(k+1-K_2)\\ &\le \frac{1}{\ln\gamma^{-1}}\cdot\frac{\gamma^{-(k+1)}}{\sqrt{k+1}} + \frac{1}{2}\int_{K_1+2}^{K_2}\frac{\gamma^{-t}}{\ln\gamma^{-1}}\cdot\frac{1}{t\sqrt{t}}dt + \frac{1}{2\ln\gamma^{-1}}\cdot\frac{\gamma^{-(k+1)}}{\sqrt{k+1}}\\ &= \frac{3}{2\ln\gamma^{-1}}\cdot\frac{\gamma^{-(k+1)}}{\sqrt{k+1}} + L(\gamma),\end{aligned} \tag{4.6}$$

where the second inequality follows from the fact that $\frac{\gamma^{-t}}{t\sqrt{t}}$ is increasing for $t > K_2$, and $L(\gamma) = \frac{1}{2}\int_{K_1+2}^{K_2}\frac{\gamma^{-t}}{\ln\gamma^{-1}}\cdot\frac{1}{t\sqrt{t}}dt$ is a constant determined by γ. Combining (4.5) and (4.6), we have

$$\begin{aligned}\sum_{j=1}^{k}\frac{\gamma^{k-j}\lambda}{\sqrt{j}} &\le \gamma^k\lambda\left(K_1\gamma^{-1} + \frac{3}{2\ln\gamma^{-1}}\cdot\frac{\gamma^{-(k+1)}}{\sqrt{k+1}} + L(\gamma)\right)\\ &= \gamma^k\lambda\left(K_1\gamma^{-1} + L(\gamma)\right) + \frac{3\lambda\gamma^{-1}}{2\ln\gamma^{-1}}\cdot\frac{1}{\sqrt{k+1}} = \mathcal{O}\left(\frac{1}{\sqrt{k}}\right)\end{aligned}$$

∎

We now present a constant steplength error bound where the steplength is fixed by a prescribed number of iterations, say K. The optimal "constant stepsize" derives the error minimizing steplength; in other words, α_k and β_k are constants for $1 \le k \le K$.

Proposition 12 (**Misspecified value iteration: constant steplength scheme**). *Suppose* $\{\tilde{v}^k\}$, $\{\widetilde{\mathbb{P}}_k\}$ *and* $\{\widetilde{\psi}_k\}$ *are generated from Algorithm 6. Suppose the learning function* $g(\cdot)$ *is strongly convex in vec*$(\mathbf{P})$ *with convexity constant* μ_g*, and is continuously differentiable in vec*$(\mathbf{P})$ *with Lipschitz gradient constant* L_g*. Suppose the learning function* $R(\cdot)$ *is strongly convex in* Ψ *with convexity constant* μ_R*, and is continuously differentiable in* Ψ *with Lipschitz gradient constant* L_R*. Suppose* $\alpha_k = \lambda_g$ *and* $\beta_k = \lambda_R$ *with* $\lambda_g > 0$ *and* $\lambda_R > 0$*. Suppose* $\mathbb{E}[\|w_k\|^2] \le \nu_g^2$ *and* $\mathbb{E}[\|u_k\|^2] \le \nu_R^2$*. Suppose* $C(s,a;\psi)$ *is Lipschitz continuous in* ψ *with constant* L_C *for all* s *and* a*. If we define* $\bar{v}^K = \frac{1}{K}\sum_{k=1}^K \tilde{v}^k$*, then*

$$\mathbb{E}[\|\bar{v}^K - v^*\|] = \mathcal{O}\left(\frac{1}{K^{1/4}}\right).$$

Proof. Instead of (4.4) in the proof of Prop. 11, we have the following

$$\mathbb{E}[\|\widetilde{\psi}_{k+1} - \psi^*\|^2] \le (1-q_g)\mathbb{E}[\|\widetilde{\psi}_k - \psi^*\|^2] + \lambda_g^2\nu_g^2,$$

where $q_g \triangleq 2\lambda_g\mu_g - \lambda_g^2 L_g^2$. Suppose λ_g is chosen such that $q_g < 1$. Thus,

$$q_g\mathbb{E}[\|\widetilde{\psi}_k - \psi^*\|^2] \le (\mathbb{E}[\|\widetilde{\psi}_k - \psi^*\|^2] - \mathbb{E}[\|\widetilde{\psi}_{k+1} - \psi^*\|^2]) + \lambda_g^2\nu_g^2.$$

Then, we have

$$\begin{aligned}\frac{1}{K}\sum_{k=1}^K q_g\mathbb{E}[\|\widetilde{\psi}_k - \psi^*\|^2] &\le \frac{1}{K}\sum_{k=1}^K (\mathbb{E}[\|\widetilde{\psi}_k - \psi^*\|^2] - \mathbb{E}[\|\widetilde{\psi}_{k+1} - \psi^*\|^2]) + \lambda_g^2\nu_g^2 \\ &\le \frac{1}{K}\mathbb{E}[\|\widetilde{\psi}_0 - \psi^*\|^2] + \lambda_g^2\nu_g^2.\end{aligned} \tag{4.7}$$

By using Hölder's inequality, Jensen's inequality applied to the counting measure, and the inequality (4.7), we have

$$\begin{aligned}\frac{1}{K}\sum_{k=1}^K \mathbb{E}[\|\widetilde{\psi}_k - \psi^*\|] &\le \frac{1}{K}\sum_{k=1}^K \sqrt{\mathbb{E}[\|\widetilde{\psi}_k - \psi^*\|^2]} \\ &\le \sqrt{\frac{1}{K}\sum_{k=1}^K \mathbb{E}[\|\widetilde{\psi}_k - \psi^*\|^2]} \\ &\le \sqrt{\frac{1}{K}\mathbb{E}[\|\widetilde{\psi}_0 - \psi^*\|^2]/q_g + \lambda_g^2\nu_g^2/q_g} \\ &\le \sqrt{\frac{a_g + \lambda_g^2\nu_g^2}{2\lambda_g\mu_g - \lambda_g^2 L_g^2}} \triangleq \sqrt{b_g(\lambda_g)},\end{aligned} \tag{4.8}$$

where $a_g = \mathbb{E}[\|\widetilde{\psi}_0 - \psi^*\|^2]/K$ and $b_g(\lambda_g) = (a_g + \lambda_g^2\nu_g^2)/(2\lambda_g\mu_g - \lambda_g^2 L_g^2)$. By taking the derivative of b_g with

respect to λ_g, we have

$$\begin{aligned}\frac{\partial b_g}{\partial \lambda_g} &= \frac{2\lambda_g\nu_g^2(2\lambda_g\mu_g - \lambda_g^2L_g^2) - (a_g + \lambda_g^2\nu_g^2)(2\mu_g - 2\lambda_gL_g^2)}{(2\lambda_g\mu_g - \lambda_g^2L_g^2)^2} \\ &= \frac{2\lambda_g^2\nu_g^2\mu_g - a_g(2\mu_g - 2\lambda_gL_g^2)}{(2\lambda_g\mu_g - \lambda_g^2L_g^2)^2} \\ &= \frac{2\nu_g^2\mu_g}{(2\lambda_g\mu_g - \lambda_g^2L_g^2)^2} \cdot \left[\lambda_g^2 + \lambda_g\frac{a_gL_g^2}{\nu_g^2\mu_g} - \frac{a_g}{\nu_g^2}\right] \\ &= \frac{2\nu_g^2\mu_g}{(2\lambda_g\mu_g - \lambda_g^2L_g^2)^2} \cdot \left[\left(\lambda_g + \frac{a_gL_g^2}{2\nu_g^2\mu_g}\right)^2 - \frac{a_g}{\nu_g^2} - \frac{a_g^2L_g^4}{4\nu_g^4\mu_g^2}\right].\end{aligned}$$

Thus, $\frac{\partial b_g}{\partial \lambda_g} = 0$ implies that $\lambda_g^* = \sqrt{\frac{a_g}{\nu_g^2} + \frac{a_g^2L_g^4}{4\nu_g^4\mu_g^2}} - \frac{a_gL_g^2}{2\nu_g^2\mu_g} = \mathcal{O}(1/\sqrt{K})$. If $0 < \lambda_g \le \lambda_g^*$, then $\frac{\partial b_g}{\partial \lambda_g} \le 0$ and thus $b_g(\lambda_g)$ is nonincreasing in λ_g; if $\lambda_g^* \le \lambda_g < \frac{2\mu_g}{\lambda_g^2}$, then $\frac{\partial b_g}{\partial \lambda_g} \ge 0$ and thus $b_g(\lambda_g)$ is nondecreasing in λ_g. Therefore, λ_g^* minimizes b_g. Then, $b_g(\lambda_g^*) = (a_g + \lambda_g^2\nu_g^2)/(2\lambda_g\mu_g - \lambda_g^2L_g^2) \le \mathcal{O}(1/\sqrt{K})$. Therefore, we have from (4.8) that

$$\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}[\|\widetilde{\psi}_k - \psi^*\|] \le \sqrt{b_g(\lambda_g)} \le \mathcal{O}\left(\frac{1}{K^{1/4}}\right). \tag{4.9}$$

Similarly, we have for suitably chosen λ_R that

$$\frac{1}{K}\sum_{k=1}^{K}\mathbb{E}[\|\mathrm{vec}(\widetilde{\mathbb{P}}_k) - \mathrm{vec}(\mathbb{P}^*)\|] \le \mathcal{O}\left(\frac{1}{K^{1/4}}\right). \tag{4.10}$$

Now, we define $\bar{v}^K = \frac{1}{K}\sum_{k=1}^{K}\tilde{v}^k$. From (4.3), we have

$$\mathbb{E}[\|\tilde{v}^{k+1} - v^*\|] \le \gamma\mathbb{E}[\|\tilde{v}^k - v^*\|] + L_C\mathbb{E}[\|\widetilde{\psi}_k - \psi^*\|] + \gamma\mathbb{E}[\|\mathrm{vec}(\widetilde{\mathbb{P}}_k) - \mathrm{vec}(\mathbb{P}^*)\|]\|v^*\|.$$

Thus,

$$(1-\gamma)\mathbb{E}[\|\tilde{v}^k - v^*\|] \le (\mathbb{E}[\|\tilde{v}^k - v^*\|] - \mathbb{E}[\|\tilde{v}^{k+1} - v^*\|]) + L_C\mathbb{E}[\|\widetilde{\psi}_k - \psi^*\|] + \gamma\mathbb{E}[\|\mathrm{vec}(\widetilde{\mathbb{P}}_k) - \mathrm{vec}(\mathbb{P}^*)\|]\|v^*\|.$$

Then, we have

$$\begin{aligned}\frac{1}{K}\sum_{k=1}^{K}(1-\gamma)\mathbb{E}[\|\tilde{v}^k - v^*\|] &\leq \frac{1}{K}\sum_{k=1}^{K}(\mathbb{E}[\|\tilde{v}^k - v^*\|] - \mathbb{E}[\|\tilde{v}^{k+1} - v^*\|]) \\ &\quad + \frac{1}{K}\sum_{k=1}^{K} L_C\mathbb{E}[\|\widetilde{\psi}_k - \psi^*\|] + \frac{1}{K}\sum_{k=1}^{K}\gamma\mathbb{E}[\|\mathrm{vec}(\widetilde{\mathbb{P}}_k) - \mathrm{vec}(\mathbb{P}^*)\|]\|v^*\| \\ &\leq \frac{1}{K}\mathbb{E}[\|\tilde{v}^0 - v^*\|] + \mathcal{O}\left(\frac{1}{K^{1/4}}\right),\end{aligned}$$

where the last inequality follows from (4.9) and (4.10). Therefore, we have

$$\begin{aligned}\mathbb{E}[\|\bar{v}^K - v^*\|] = \mathbb{E}\left[\|\frac{1}{K}\sum_{k=1}^{K}\tilde{v}^k - v^*\|\right] = \mathbb{E}\left[\|\frac{1}{K}\sum_{k=1}^{K}(\tilde{v}^k - v^*)\|\right] &\leq \frac{1}{K}\sum_{k=1}^{K}\mathbb{E}[\|\tilde{v}^k - v^*\|] \\ \leq \frac{\mathbb{E}[\|\tilde{v}^0 - v^*\|]/(1-\gamma)}{K} + \mathcal{O}\left(\frac{1}{K^{1/4}}\right).&\end{aligned}$$

∎

4.3 Misspecified policy iteration

In this section, we consider a policy iteration scheme for the resolution of misspecified MDPs. We initiate our discussion with a formal statement of the misspecified policy iteration scheme and subsequently prove its asymptotic convergence. If $c^\pi(\cdot) \triangleq C(\cdot, \pi(\cdot); \psi^*)$ and $\tilde{c}^{\pi_k}(\cdot) \triangleq C(\cdot, \pi_k(\cdot); \tilde{\psi}_k)$, then the operators $\mathcal{M}^\pi$ and $\widetilde{\mathcal{M}}_k^{\pi_k}$ may be defined as follows for policies π and π_k, respectively:

$$\mathcal{M}^\pi v \triangleq c^\pi + \gamma(\mathbb{P}^*)^\pi v \text{ and } \widetilde{\mathcal{M}}_k^{\pi_k} v \triangleq \tilde{c}^{\pi_k} + \gamma\widetilde{\mathbb{P}}_k^{\pi_k} v,$$

Next, we define the misspecified policy iteration scheme.

Algorithm 7 (**Misspecified policy Iteration**). **Step 0:** *Let* $\tilde{v}^0 : \mathcal{S} \to \mathbb{R}$, $vec(\widetilde{\mathbb{P}}_0) \in vec(\mathbf{P})$, $\alpha_k > 0$, $\widetilde{\psi}_0 \in \Psi$, $\alpha_0 > 0$, $\beta_0 > 0$ *and* $k = 0$.

Step 1: *For all* $k \geq 0$,

$$a_{k+1}(s) := \underset{a \in \mathcal{A}}{\operatorname{argmax}}(C(s, a; \widetilde{\psi}_k) + \gamma \widetilde{\mathbb{P}}_k^{\pi_{k+1}} \tilde{v}^{k+1}), \tag{Computation}$$

$$vec(\widetilde{\mathbb{P}}_{k+1}) := \Pi_{vec(\mathbf{P})} \left(vec(\widetilde{\mathbb{P}}_k) - \alpha_k (\nabla g(\widetilde{\mathbb{P}}_k) + w_k) \right), \tag{Learning}-\mathbb{P}$$

$$\widetilde{\psi}_{k+1} := \Pi_{\Psi} \left(\widetilde{\psi}_k - \beta_k (\nabla R(\widetilde{\psi}_k) + u_k) \right), \tag{Learning}-\Psi$$

where $w_k \triangleq \nabla g(\widetilde{\mathbb{P}}_k; \eta_k) - \nabla g(\widetilde{\mathbb{P}}_k)$ *with* $g(\mathbb{P}) \triangleq \mathbb{E}[g(\mathbb{P}; \eta)]$, $u_k = \nabla R(\widetilde{\psi}_k; \xi_k) - \nabla R(\widetilde{\psi}_k)$, $R(\psi) \triangleq \mathbb{E}[R(\psi; \xi)]$ *and* $(I - \gamma \widetilde{\mathbb{P}}_k^{\pi_k}) \tilde{v}^{k+1} = \tilde{c}^{\pi_k}$.

Step 2: *If* $k > K$, *stop; else* $k := k + 1$ *and go to Step 1.*

We now provide a lemma that provides the error bound for the approximate policy iteration, which is useful for our rate analysis.

Lemma 14 (Approximate policy iteration bound (cf. p.48 in [105]))**.** *Let* $\tilde{v}^k$ *be the approximate value function. Suppose for all* k

$$\|v^k - \tilde{v}^k\| \leq \delta,$$

and

$$\|\mathcal{M}^{\pi^*} \tilde{v}^k - \mathcal{M}^{\pi_{k+1}} \tilde{v}^k\| \leq \epsilon.$$

Then, we have

$$\limsup_{k \to \infty} \|v^{k+1} - v^*\| \leq \frac{\epsilon + 2\gamma\delta}{(1-\gamma)^2}.$$

Analogous to Proposition 11 for the value iteration, we can get the following convergence statement where $\| \bullet \|$ denotes the infinity norm for both matrices and vectors.

Proposition 13 (**Misspecified policy iteration: a.s. convergence and rate statement**). *Suppose* $\{\tilde{v}^k\}$, $\{\widetilde{\mathbb{P}}_k\}$ *and* $\{\widetilde{\psi}_k\}$ *are generated by Algorithm 7 and the learning functions* $g(\cdot)$ *and* $R(\cdot)$ *are strongly convex. Finally, suppose* $C(s,a;\psi)$ *is Lipschitz continuous in* ψ *with constant* L_C *for all* s *and* a *and* $\|\tilde{v}^k\|$ *is bounded a.s. for all* k. *Suppose* $\|c^{\pi^*} - c^{\pi_k}\| \leq \Delta$ *and* $\|(\mathbb{P}^*)^{\pi^*} - (\mathbb{P}^*)^{\pi_k}\| \leq \Delta$ *for all* k. *Then, the following hold:*

(i) $\|\tilde{v}^k - v^*\| \to 0 \quad a.s. \quad as\ k \to \infty$.

(ii) For any k, *we have that the following holds:*

$$\mathbb{E}[\|\tilde{v}^{k+1} - v^*\|] = \mathcal{O}\left(\frac{(1+\gamma)\Delta}{(1-\gamma)^2}\right) + \mathcal{O}\left(\frac{(1+\gamma^2)(1+\gamma)}{(1-\gamma)^3\sqrt{k}}\right).$$

Proof. **(i)** We proceed to show that $\|v^k - \tilde{v}^k\| \to 0$ as $k \to \infty$ whereby the result follows by recalling that by the convergence of policy iteration, $\|v^k - v^*\| \to 0$ as $k \to \infty$. From Algorithm 7, we have

$$\begin{aligned}
\|v^{k+1} - \tilde{v}^{k+1}\| &= \|c^{\pi_k} + \gamma(\mathbb{P}^*)^{\pi_k} v^{k+1} - (\tilde{c}^{\pi_k} + \gamma\widetilde{\mathbb{P}}_k^{\pi_k}\tilde{v}^{k+1})\| \\
&= \|c^{\pi_k} - \tilde{c}^{\pi_k} + \gamma(\mathbb{P}^*)^{\pi_k}(v^{k+1} - \tilde{v}^{k+1}) + \gamma((\mathbb{P}^*)^{\pi_k} - \widetilde{\mathbb{P}}_k^{\pi_k})\tilde{v}^{k+1}\| \\
&\leq L_C N\|\psi^* - \widetilde{\psi}_k\| + \gamma\|(\mathbb{P}^*)^{\pi_k}\|\|v^{k+1} - \tilde{v}^{k+1}\| + \gamma\|(\mathbb{P}^*)^{\pi_k} - \widetilde{\mathbb{P}}_k^{\pi_k}\|\|\tilde{v}^{k+1}\|.
\end{aligned}$$

It follows that

$$\begin{aligned}
\|v^{k+1} - \tilde{v}^{k+1}\| &\leq \frac{L_C N\|\psi^* - \widetilde{\psi}_k\| + \gamma\|(\mathbb{P}^*)^{\pi_k} - \widetilde{\mathbb{P}}_k^{\pi_k}\|\|\tilde{v}^{k+1}\|}{1 - \gamma\|(\mathbb{P}^*)^{\pi_k}\|} \\
&= \frac{L_C N\|\psi^* - \widetilde{\psi}_k\| + \gamma\|(\mathbb{P}^*)^{\pi_k} - \widetilde{\mathbb{P}}_k^{\pi_k}\|\|\tilde{v}^{k+1}\|}{1 - \gamma}.
\end{aligned} \tag{4.11}$$

Recall that the learning problem for ψ^* and $\mathbb{P}^*$ are both strongly convex, implying that $\widetilde{\psi}_k \to \psi^*$ and $\text{vec}(\widetilde{\mathbb{P}}_k) \to \text{vec}(\mathbb{P}^*)$ *a.s.* as $k \to \infty$. Thus, by the *a.s.* boundedness of $\tilde{v}^k$ and by invoking the property that $\|v^k - v^*\| \to 0$ as $k \to \infty$, we have that $\|v^k - \tilde{v}^k\| \to 0$ *a.s.* as $k \to \infty$. Therefore, $\|\tilde{v}^k - v^*\| \to 0$ *a.s.* as $k \to \infty$.

(ii) By taking expectations on both sides of (4.11), we have the following:

$$\mathbb{E}[\|v^{k+1} - \tilde{v}^{k+1}\|] \leq \frac{L_C N\mathbb{E}[\|\psi^* - \widetilde{\psi}_k\|] + \gamma\mathbb{E}[\|(\mathbb{P}^*)^{\pi_k} - \widetilde{\mathbb{P}}_k^{\pi_k}\|\|\tilde{v}^{k+1}\|]}{1 - \gamma}. \tag{4.12}$$

Recall that the learning problem for ψ^* and $\mathbb{P}^*$ are both strongly convex. Then, we can use the standard

rate estimate (see (5.292) in [42]) to get the following:

$$\mathbb{E}[\|\widetilde{\psi}_k - \psi^*\|] = \mathcal{O}\left(\frac{1}{\sqrt{k}}\right) \text{ and } \mathbb{E}[\|(\mathbb{P}^*)^{\pi_k} - \widetilde{\mathbb{P}}_k^{\pi_k}\|] = \mathcal{O}\left(\frac{1}{\sqrt{k}}\right). \tag{4.13}$$

Consequently, we obtain the following:

$$\mathbb{E}[\|v^{k+1} - \tilde{v}^{k+1}\|] = \frac{1+\gamma}{1-\gamma}\mathcal{O}\left(\frac{1}{\sqrt{k}}\right).$$

On the other hand,

$$\mathbb{E}[\|\mathcal{M}^{\pi^*}\tilde{v}^k - \mathcal{M}^{\pi_{k+1}}\tilde{v}^k\|] = \mathbb{E}[\|c^{\pi^*} + \gamma(\mathbb{P}^*)^{\pi^*}\tilde{v}^k - (c^{\pi_{k+1}} + \gamma(\mathbb{P}^*)^{\pi_{k+1}}\tilde{v}^k)\|] = (1+\gamma)\mathcal{O}(\Delta).$$

Then, we may use the approximate policy iteration bound (Lemma 14) to get the following:

$$\mathbb{E}[\|v^{k+1} - v^*\|] = \frac{(1+\gamma)\mathcal{O}(\Delta) + 2\gamma \cdot \frac{1+\gamma}{1-\gamma}\mathcal{O}\left(\frac{1}{\sqrt{k}}\right)}{(1-\gamma)^2}.$$

Therefore,

$$\mathbb{E}[\|\tilde{v}^{k+1} - v^*\|] = \mathcal{O}\left(\frac{(1+\gamma)\Delta}{(1-\gamma)^2}\right) + \mathcal{O}\left(\frac{(1+\gamma^2)(1+\gamma)}{(1-\gamma)^3\sqrt{k}}\right).$$

∎

4.4 Misspecified Q-learning

When transition matrices are unavailable, a commonly adopted approach is a simulated approach popularly referred to as Q-learning [97]. We consider a misspecified variant of Q-learning that incorporates learning of the misspecified cost and examines the resulting sequence of estimators. We begin by defining the Q-function as $Q(s,a) \triangleq C(s,a;\psi^*) + \gamma \sum_{s' \in \mathcal{S}} \mathbb{P}^*(s'|s,a)v(s')$, which allows for restating as follows:

$$Q(s,a) \triangleq C(s,a;\psi^*) + \gamma \sum_{s' \in \mathcal{S}} \mathbb{P}^*(s'|s,a) \max_{b \in \mathcal{A}} Q(s',b). \tag{4.14}$$

We define the operator $\mathcal{T}$ as

$$\mathcal{T}[Q(s,a)] \triangleq C(s,a;\psi^*) + \gamma \sum_{s' \in \mathcal{S}} \mathbb{P}^*(s'|s,a) \max_{b \in \mathcal{A}} Q(s',b).$$

Then the Q-function is the fixed point of the operator $\mathcal{T}$; i.e. $Q = \mathcal{T}[Q]$. Given the vector $\widetilde{\psi}_k$ in the cost at iteration n, we may define the misspecified operator $\widetilde{\mathcal{T}}_k$ at iteration n as

$$\widetilde{\mathcal{T}}_k Q(s,a) \triangleq C(s,a;\widetilde{\psi}_k) + \gamma \sum_{s' \in \mathcal{S}} \mathbb{P}^*(s'|s,a) \max_{b \in \mathcal{A}} Q(s',b),$$

where $\widetilde{\mathcal{T}}_k Q$ is used to denote $\widetilde{\mathcal{T}}_k[Q]$. As in previous sections, we may specify our misspecified Q-learning scheme as follows:

Algorithm 8 (**Misspecified Q-learning**). **Step 0:** *Let $\tilde{Q}_0(s,a) \in \mathbb{R}$, $\widetilde{\psi}_0 \in \Psi$, $\beta_0 > 0$ and $k = 0$.*

Step 1: *For all $n \geq 0$,*

$$\widetilde{Q}_{k+1}(s,a) := (1-\delta)\widetilde{Q}_k(s,a) + \delta \left[C(s,a;\widetilde{\psi}_k) + \gamma \max_{b \in \mathcal{A}} \widetilde{Q}_k(s',b) \right], \qquad (Q-\text{update})$$

$$\widetilde{\psi}_{k+1} := \Pi_\Psi \left(\widetilde{\psi}_k - \beta_k (\nabla R(\widetilde{\psi}_k) + u_k) \right), \qquad (\text{Learning-}\psi)$$

where $\delta \in (0,1)$, s' is the random next state reached when the current state is s and action is a, and $u_k = \nabla R(\widetilde{\psi}_k;\xi_k) - \nabla R(\widetilde{\psi}_k)$ with $R(\psi) \triangleq \mathbb{E}[R(\psi;\xi)]$.

Step 2: *If $n > K$, stop; else $k := k+1$ and go to Step 1.*

Our convergence analysis begins with a reproduction of two classical results regarding the operator $\widetilde{\mathcal{T}}$, which may be directly applied to the misspecified operator $\widetilde{\mathcal{T}}_k$. First, $\widetilde{\mathcal{T}}_k$ is a contraction mapping.

Proposition 14 (**Contractive property of $\widetilde{\mathcal{T}}_k$** [98]). *If $0 \leq \gamma < 1$, then $\|\widetilde{\mathcal{T}}_k[Q_1] - \widetilde{\mathcal{T}}_k[Q_2]\|_\infty \leq \gamma \|Q_1 - Q_2\|_\infty$ for any two vectors Q_1 and Q_2.* ∎

Second, the estimated Q-function stays bounded.

Proposition 15 (**Boundedness of Q function** [106]). *There exists $\hat{Q}_{\max}$ such that $\|\hat{Q}_k\|_\infty \leq \hat{Q}_{\max}$ for any k.* ∎

We now provide an intermediate lemma that provides a constant steplength error bound on a suitably defined metric D.

Lemma 15. *For any state-action pair (s,a), suppose $D_k(s,a) = \overline{Q}_k(s,a) - z_k$ and $z_k(s,a)$ be defined as follows:*

$$z_{k+1}(s,a) = (1-\delta) z_k(s,a) + \delta\gamma w_k(s,a), \quad z_0(s,a) = 0. \qquad (4.15)$$

Then for any k, we have that $\mathbb{E}[\|D_k\|_\infty] \leq \left(\mathcal{O}\left(\frac{1}{\sqrt{k}}\right) + \frac{\gamma^2}{1-\gamma} \sqrt{\frac{\delta W_{\max}^2}{2-\delta}} \right)$.

Proof. We utilize an approach employed in [107] and begin by defining the error $\overline{Q}_k(s,a)$ as $\overline{Q}_k(s,a) \triangleq \widetilde{Q}_k(s,a) - Q(s,a)$. Using (4.14) and ($Q-$update), the error can be written as

$$\begin{aligned}
\overline{Q}_{k+1}(s,a) &= (1-\delta)\overline{Q}_k(s,a) + \delta\left[C(s,a;\widetilde{\psi}_k) + \gamma \max_{b\in\mathcal{A}} \widetilde{Q}_k(s',b) - Q(s,a)\right] \\
&= (1-\delta)\overline{Q}_k(s,a) + \delta\left[C(s,a;\widetilde{\psi}_k) - C(s,a;\psi^*) + \gamma \max_{b\in\mathcal{A}} \widetilde{Q}_k(s',b) - \gamma \sum_{s'\in\mathcal{S}} \mathbb{P}^*(s'|s,a) \max_{b\in\mathcal{A}} Q(s',b)\right] \\
&= (1-\delta)\overline{Q}_k(s,a) + \delta\left(C(s,a;\widetilde{\psi}_k) - C(s,a;\psi^*)\right) \\
&+ \delta\gamma \sum_{s'\in\mathcal{S}} \mathbb{P}^*(s'|s,a)\left(\max_{b\in\mathcal{A}} \widetilde{Q}_k(s',b) - \max_{b\in\mathcal{A}} Q(s',b)\right) + \delta\gamma w_k(s,a) \\
&= (1-\delta)\overline{Q}_k(s,a) + \delta\left(C(s,a;\widetilde{\psi}_k) - C(s,a;\psi^*)\right) + \delta(T\widetilde{Q}_k(s,a) - TQ(s,a)) + \delta\gamma w_n(s,a),
\end{aligned}$$

where $w_k(s,a) = \max_{b\in\mathcal{A}} \widetilde{Q}_k(s',b) - \sum_{s'\in\mathcal{S}} \mathbb{P}^*(s'|s,a) \max_{b\in\mathcal{A}} \widetilde{Q}_n(s',b)$. If z_k is defined by (4.15) (as done in [107]), then the following holds for the second moment:

$$\mathbb{E}[\|z_k\|_2] \leq \sqrt{\frac{\gamma^2\delta W_{\max}^2}{2-\delta}}, \tag{4.16}$$

where $W_{\max}^2 = |\mathcal{S}\times\mathcal{A}|4\widehat{Q}_{\max}^2$ with $|\mathcal{S}|$ being the cardinality of the set of states and $|\mathcal{A}|$ being the cardinality of the set of possible actions. By defining the sequence $D_k \triangleq \overline{Q}_k - z_k$, we may bound it as follows:

$$\begin{aligned}
& D_{k+1}(s,a) = (1-\delta)D_k(s,a) + \delta\left(C(s,a;\widetilde{\psi}_k) - C(s,a;\psi^*)\right) + \delta(T\widetilde{Q}_k(s,a) - TQ(s,a)) \\
\implies \quad & |D_{k+1}(s,a)| \leq (1-\delta)|D_k(s,a)| + \delta L_C\|\widetilde{\psi}_k - \psi^*\| + \delta\|T\widetilde{Q}_k(s,a) - TQ(s,a)\|_\infty \\
& \leq (1-\delta)|D_k(s,a)| + \delta L_C\|\widetilde{\psi}_k - \psi^*\| + \delta\gamma\|\widetilde{Q}_k - Q\|_\infty \\
& \leq (1-\delta)\|D_k\|_\infty + \delta L_C\|\widetilde{\psi}_k - \psi^*\| + \delta\gamma\|\overline{Q}_k\|_\infty,
\end{aligned}$$

where the first inequality follows from the Lipschitz continuity of the cost function and the second inequality follows from Proposition 14. Therefore,

$$\begin{aligned}
\|D_{k+1}\|_\infty &\leq (1-\delta)\|D_k\|_\infty + \delta L_C\|\widetilde{\psi}_k - \psi^*\| + \delta\gamma\|\overline{Q}_k\|_\infty \\
&\leq (1-\delta)\|D_k\|_\infty + \delta L_C\|\widetilde{\psi}_k - \psi^*\| + \delta\gamma(\|D_k\|_\infty + \|z_n\|_\infty) \\
&= (1-\delta(1-\gamma))\|D_k\|_\infty + \delta L_C\|\widetilde{\psi}_k - \psi^*\| + \delta\gamma\|z_k\|_\infty.
\end{aligned}$$

We may then derive a bound for D_k:

$$\begin{aligned}
\|D_k\|_\infty &\le (1-\delta(1-\gamma))\|D_{k-1}\|_\infty + \delta L_C\|\widetilde{\psi}_{k-1}-\psi^*\| + \delta\gamma\|z_{k-1}\|_\infty \\
&\le (1-\delta(1-\gamma))^2\|D_{k-2}\|_\infty + (1-\delta(1-\gamma))\delta L_C\|\widetilde{\psi}_{k-2}-\psi^*\| + \delta L_C\|\widetilde{\psi}_{k-1}-\psi^*\| \\
&\quad + (1-\delta(1-\gamma))\delta\gamma\|z_{k-2}\|_\infty + \delta\gamma\|z_{k-1}\|_\infty \\
&\le \quad \vdots \\
&\le (1-\delta(1-\gamma))^k\|D_0\|_\infty + \delta L_C\sum_{l=0}^{k-1}(1-\delta(1-\gamma))^l\|\widetilde{\psi}_{k-1-l}-\psi^*\| + \delta\gamma\sum_{l=0}^{k-1}(1-\delta(1-\gamma))^l\|z_{k-1-l}\|_\infty.
\end{aligned}$$

Recall that the learning problem for ψ^* is strongly convex implying that for some λ and for all k, we have $\mathbb{E}[\|\widetilde{\psi}_k-\psi^*\|] \le \frac{\lambda}{\sqrt{k}}$. Therefore,

$$\begin{aligned}
\mathbb{E}[\|D_k\|_\infty] &\le (1-\delta(1-\gamma))^k\|\overline{Q}_0\|_\infty + \delta L_C\sum_{l=0}^{k-1}\frac{(1-\delta(1-\gamma))^l\lambda}{\sqrt{k-1-l}} + \delta\gamma\sum_{l=0}^{k-1}(1-\delta(1-\gamma))^l\|z_{k-1-l}\|_\infty \\
&\le (1-\delta(1-\gamma))^k\|\overline{Q}_0\|_\infty + \delta L_C\sum_{l=0}^{k-1}\frac{(1-\delta(1-\gamma))^l\lambda}{\sqrt{k-1-l}} + \frac{\delta\gamma}{\delta(1-\gamma)}\sqrt{\frac{\gamma^2\delta W_{\max}^2}{2-\delta}} \qquad (4.17) \\
&= \mathcal{O}\left(\frac{1}{\sqrt{k}}\right) + \frac{\gamma^2}{1-\gamma}\sqrt{\frac{\delta W_{\max}^2}{2-\delta}},
\end{aligned}$$

where the second inequality utilizes $\mathbb{E}[\|z_k\|_\infty] \le \mathbb{E}[\|z_k\|_2]$ together with the bound (4.16) and the last equality utilizes a proof technique similar to that adopted in Prop. 11. ∎

Proposition 16 (Constant steplength error bound for misspecified Q-learning). *Suppose $\{\widetilde{Q}_k\}$, and $\{\widetilde{\psi}_k\}$ are generated from Algorithm 8. Suppose the learning function $R(\cdot)$ is strongly convex in Ψ and $C(s,a;\psi)$ is Lipschitz continuous in ψ with constant L_C for all s and a. Then, the following holds for any k and $\delta < 1$:*

$$\mathbb{E}\left[\|\overline{Q}_k\|_\infty\right] \le \mathcal{O}\left(\frac{1}{\sqrt{k}}\right) + \frac{\gamma}{1-\gamma}\sqrt{\frac{\delta W_{\max}^2}{2-\delta}}.$$

Proof. The result follows directly from Lemma 15, expression (4.16), and $\delta < 1$:

$$\begin{aligned}
\mathbb{E}\left[\|\overline{Q}_k\|_\infty\right] &\le \mathcal{O}\left(\frac{1}{\sqrt{k}}\right) + \frac{\gamma^2}{1-\gamma}\sqrt{\frac{\delta W_{\max}^2}{2-\delta}} + \sqrt{\frac{\gamma^2\delta W_{\max}^2}{2-\delta}} \\
&= \mathcal{O}\left(\frac{1}{\sqrt{k}}\right) + \frac{\gamma}{1-\gamma}\sqrt{\frac{\delta W_{\max}^2}{2-\delta}} \\
&= \mathcal{O}\left(\frac{1}{\sqrt{k}}\right) + \mathcal{O}\left(\sqrt{\delta}\right).
\end{aligned}$$

∎

Suppose we take m learning steps in ψ before updating the Q function. Then, we may specify our misspecified Q-learning scheme as follows:

Algorithm 9 (**Misspecified Q-learning with multiple steps of learning**). **Step 0:** *Let $\widetilde{Q}_0(s,a) \in \mathbb{R}$, $\widetilde{\psi}_0^{(0)} \in \Psi$, $\beta_0 > 0$ and $k = 0$.*

Step 1: *For all $n \geq 0$,*

$$\widetilde{Q}_{k+1}(s,a) := (1-\delta)\widetilde{Q}_k(s,a) + \delta\left[C(s,a;\widetilde{\psi}_k^{(m)}) + \gamma \max_{b\in\mathcal{A}} \widetilde{Q}_k(s',b)\right], \qquad (Q-\text{update})$$

$$\widetilde{\psi}_{k+1}^{(0)} := \widetilde{\psi}_k^{(m)}, \qquad (\text{Learning-}\psi)$$

$$\widetilde{\psi}_k^{(l)} := \Pi_\Psi\left(\widetilde{\psi}_k^{(l-1)} - \beta_k^{(l-1)}(\nabla R(\widetilde{\psi}_k^{(l-1)}) + u_k^{(l-1)})\right), \quad l = 1,\ldots,m,$$

where $\delta \in (0,1)$, s' is the random next state reached when the current state is s and action is a, and $u_k^{(l)} = \nabla R(\widetilde{\psi}_k^{(l)};\xi_k^{(l)}) - \nabla R(\widetilde{\psi}_k^{(l)})$ with $R(\psi) \triangleq \mathbb{E}[R(\psi;\xi)]$.

Step 2: *If $n > K$, stop; else $k := k+1$ and go to Step 1.*

Proposition 17 (**Constant steplength error bound for misspecified Q-learning with multiple steps of learning**). *Suppose $\{\widetilde{Q}_k\}$, and $\{\widetilde{\psi}_k^{(l)}\}$ are generated from Algorithm 9. Suppose the learning function $R(\cdot)$ is strongly convex in Ψ and $C(s,a;\psi)$ is Lipschitz continuous in ψ with constant L_C for all s and a. Then, the following holds for any k and $\delta < 1$:*

$$\mathbb{E}\left[\|\overline{Q}_k\|_\infty\right] \leq \mathcal{O}\left(\frac{1}{\sqrt{mk}}\right) + \frac{\gamma}{1-\gamma}\sqrt{\frac{\delta W_{\max}^2}{2-\delta}}.$$

Proof. Recall that the learning problem for ψ^* is strongly convex implying that for some λ and for all k, we have $\mathbb{E}[\|\widetilde{\psi}_k^{(m)} - \psi^*\|] \leq \frac{\lambda}{\sqrt{mk}} = \frac{\lambda/\sqrt{m}}{\sqrt{k}}$. By using the same technique in Lemma 15, we have a similar bound for $\mathbb{E}[\|D_k\|_\infty]$ in (4.17):

$$\begin{aligned}
\mathbb{E}[\|D_k\|_\infty] &\leq (1-\delta(1-\gamma))^k\|\overline{Q}_0\|_\infty + \delta L_C\sum_{l=0}^{k-1}\frac{(1-\delta(1-\gamma))^l\lambda}{\sqrt{m}\sqrt{k-1-l}} + \delta\gamma\sum_{l=0}^{k-1}(1-\delta(1-\gamma))^l\|z_{k-1-l}\|_\infty \\
&\leq (1-\delta(1-\gamma))^k\|\overline{Q}_0\|_\infty + \delta L_C\sum_{l=0}^{k-1}\frac{(1-\delta(1-\gamma))^l\lambda}{\sqrt{m}\sqrt{k-1-l}} + \frac{\delta\gamma}{\delta(1-\gamma)}\sqrt{\frac{\gamma^2\delta W_{\max}^2}{2-\delta}} \qquad (4.18)\\
&= \mathcal{O}\left(\frac{1}{\sqrt{mk}}\right) + \frac{\gamma^2}{1-\gamma}\sqrt{\frac{\delta W_{\max}^2}{2-\delta}},
\end{aligned}$$

where the second inequality utilizes $\mathbb{E}[\|z_k\|_\infty] \leq \mathbb{E}[\|z_k\|_2]$ together with the bound (4.16) and the last equality utilizes a proof technique similar to that adopted in Prop. 11. Then, the result follows directly from (4.18),

expression (4.16), and $\delta < 1$:

$$\mathbb{E}\left[\|\overline{Q}_k\|_\infty\right] \leq \mathcal{O}\left(\frac{1}{\sqrt{mk}}\right) + \frac{\gamma^2}{1-\gamma}\sqrt{\frac{\delta W_{\max}^2}{2-\delta}} + \sqrt{\frac{\gamma^2 \delta W_{\max}^2}{2-\delta}} = \mathcal{O}\Big(\frac{1}{\sqrt{mk}}\Big) + \mathcal{O}\Big(\sqrt{\delta}\Big).$$

∎

4.5 Numerical results

4.5.1 Problem setting

We consider a Markov decision problem. There is a chain of N states, which are labeled consecutively from left to right, $s = 1, 2, \ldots, N$. An agent has two possible actions, go to the left (lower state numbers; $a = -1$), or go to the right (higher state numbers; $a = +1$). Both the first and last states in the chain, states number 1 and N, are rewarded with $r(1) = r(N) = 1$. The reward of the intermediate states is set to a small negative value, i.e. $r(i) = -0.1$ for $1 < i < N$. We consider a discount factor $\gamma = 0.9$.

If the agent wants to move to the left ($a = -1$), with probability $P_1 = 0.8$ the system responds with the correct move from the intended. So, the agent will move to the right with probability $1 - P_1 = 0.2$. Similarly, if the agent wants to move to the right ($a = 1$), the system responds with the correct move from the intended with probability $P_2 = 0.6$. The transition probabilities $T(s'|s, a)$ for this example are zero expect for the following elements,

$$\begin{aligned}
&T(1|1, \pm 1) = 1, && T(N|N, \pm 1) = 1,\\
&T(s-1|s, 1) = 1 - P_2, && T(s+1|s, 1) = P_2, \quad 1 < s < N,\\
&T(s-1|s, -1) = P_1, && T(s+1|s, -1) = 1 - P_1, \quad 1 < s < N.
\end{aligned}$$

The first two entries specify the ends of the chain as absorbing boundaries as the agent would stay in this state once it reaches these states.

For learning the reward function, we first generate L N-dimensional random vectors $X_i \in \mathbb{R}^N$, $i = 1, \ldots, L$, such that $X_i(s)$ is a normal random variables with mean $r(s)$ and variance $r(s)^2/4$ for $1 < s < N$, $i = 1, \ldots, L$ with $L = 1000$. We assume that $r(s) \equiv r$ for $1 < s < N$. Our estimator for r is $\hat{r}$, which solves the following optimization problems:

$$\min_{\hat{r}} \mathbb{E}\left[\sum_{s=2}^{N-1}\left(\hat{r} - \frac{\sum_{i=1}^{L} X_i(s)}{L}\right)^2\right]. \tag{4.19}$$

For learning the transition matrices, we first generate two N-dimensional random vectors $Y_1 \in \mathbb{R}^N$ and $Y_2 \in \mathbb{R}^N$, such that $Y_1(s)$ is a binomial random variables with parameters $L = 1000$ and $P_1 = 0.8$ for $1 < s < N$, and $Y_2(s)$ ia a binomial random variables with parameters $L = 1000$ and $P_2 = 0.6$ for $1 < s < N$. Our estimators for P_1 and P_2 are $\hat{P}_1$ and $\hat{P}_2$, respectively, which solve the following optimization problems:

$$\min_{\hat{P}_i} \mathbb{E}\left[\sum_{s=2}^{N-1}\left(\hat{P}_i - \frac{Y_i(s)}{L}\right)^2\right], \tag{4.20}$$

for $i = 1, 2$.

4.5.2 Results

We use the value iteration to generate 15 sample paths for each dimension of the transition matrices. We stop at $k = 1000$. If we use constant steplength $\alpha_k = \beta_k = 0.01$ for the learning problem in the value iteration, we can get

Table 4.1: Misspecified value iteration

N	$\mathbb{E}[\|\tilde{v}^k - v^*\|/\|v^*\|]$
10	4.0×10^{-3}
20	6.2×10^{-3}
50	3.9×10^{-3}
100	3.1×10^{-3}

Next, we use the policy iteration for each dimension of the transition matrices. We stop when $\|\tilde{v}^{k+1} - \tilde{v}^k\| < 10^{-4}$. If we use constant steplength $\alpha_k = \beta_k = 0.01$ for the learning problem in the policy iteration, we can get

Table 4.2: Misspecified policy iteration

N	$\|\tilde{v}^k - v^*\|/\|v^*\|$	Number of iteration	Number of iteration given $\mathbb{P}^*$ and r
10	2.3×10^{-3}	18	3
20	9.3×10^{-3}	7	4
50	2.7×10^{-3}	7	5
100	4.4×10^{-3}	8	4

Finally, we use Q-learning for each dimension of the transition matrices. We stop when $\|\tilde{Q}^{k+1} - \tilde{Q}^k\| < 10^{-4}$. If we use constant steplength $\beta_k = 0.01$ for the learning problem, we can get

Table 4.3: Misspecified Q-learning

N	Number of iteration	Number of iteration given r
10	135	125
20	85	55
50	389	919
100	769	760

4.6 Concluding remarks

Motivated by the increasing role of streaming data and misspecification in decision-making problems, we consider the resolution of MDPs in which transition matrices are unknown and the cost functions are misspecified. We develop extensions to value iteration, policy iteration and Q-learning through which both misspecification is resolved while solving the original MDP in an asymptotic sense. A precise characterization of the impact of learning on the resulting error bounds is provided in the context of value iteration and Q-learning.

We conclude with a short commentary on the nature of the error bounds. First, we assume that the learning problems are strongly convex since deriving overall rate statements requires bounds on the expected error in parameter estimates. In fact, the knowledge of the convexity constant in the learning problem is crucial in the development of bounds. It is worth emphasizing that if mere convexity assumptions are imposed on the learning problems, the currently adopted avenue cannot be utilized since error bounds are only available in a functional value sense. Furthermore, while averaging-based techniques may be utilized to resolve merely convex learning problems, such approaches provide bounds on the averaged iterates in a functional sense but not on the solution iterates; in the absence of bounds on the solution iterates, one cannot derive rate statements. Second, in the context of Q-learning, we develop a misspecified variant of the constant steplength scheme. Naturally, diminishing steplength versions can also be developed which will be the subject of future work. Third, throughout the chapter, we assume that the learning problems are static and consequently, rather than regret-based bounds, we derive error bounds on the optimal functional value or solution.

Chapter 5

Conclusions

In this thesis, we consider a broad class of computational problems that have historically been addressed in a regime when their parameters are known a priori. Yet, as we contend with challenges posed by the presence of streaming data, growing uncertainty, and informational inadequacy, we can no longer work under the premise that such parameters are available. Instances of such parameters include the covariance matrix in a portfolio optimization problem, distributional parameters of arrival and service processes in a queueing system, and machine efficiencies in a production network. In many instances, these parameters may be estimated through a separate learning problem. In fact, the traditional approach has been to first learn such parameters and subsequently solve the parametrized computational problem. However, if the learning problem is a stochastic optimization problem, resolving the learning problem may require simulation-based schemes and provide exact solutions only in a limiting sense. In practical settings, the learning process has to be terminated finitely and thus leading to an erroneous estimator of the parameter which in turn leads to the error cascading into the solution of the subsequent computational problem. We pursue a rather different tack that solves the learning and computational problem *simultaneously* rather than sequentially and consider three types of computational problems: (i) Misspecified stochastic optimization and variational inequality problems; (ii) Misspecified stochastic Nash games; nd (iii) misspecified Markov decision processes.

We first consider a misspecified stochastic optimization problem in which the objective is parameterized by a vector that can be learnt by solving a suitably defined learning problem. In both strongly convex and merely convex regimes, we develop a set of coupled stochastic approximation schemes which produces schemes such that almost sure convergence can be guaranteed for both the solution and parameters. Error bounds are also provided for both regimes. For strongly convex problems, the optimal rate of convergence is recovered while in convex regimes there is a degradation in error, i.e. $\mathcal{O}\left(\frac{\sqrt{\ln(K)}}{\sqrt{K}}\right)$ instead of $\mathcal{O}\left(\frac{1}{\sqrt{K}}\right)$. When the averaging window is modified suitably, it can be seen that the original rate of of $\mathcal{O}\left(\frac{1}{\sqrt{K}}\right)$ is recovered. Also, we can get an error bound for the average regret in the online decision-making setting, i.e. $\mathcal{O}\left(\frac{\ln K}{\sqrt{K}}\right)$ for a suitably chosen steplength. As the generalization of the misspecified optimization problem, a misspecified stochastic variational inequality problem is considered, and we propose analogous stochastic approximation

schemes for simultaneous computation and learning. Almost-sure convergence statements and error analysis can be provided. Specially, for merely monotone maps, we employ (Tikhonov) regularized scheme, and we can quantify the degradation associated with learning under suitable weak-sharpness assumption.

We then consider misspecified Nash games and present schemes for learning equilibria and parameters under two settings. First, we consider convex stochastic Nash games in which agent payoffs are parameterized by a misspecified vector. We propose schemes that combine a gradient step and a stochastic approximation step. Equilibria and the true parameters can be both shown to be achieved in an almost sure sense. Second, we consider stochastic Nash-Cournot games where we assume common knowledge holds but aggregate output is unobservable. In such a setting, we propose an iterative fixed-point scheme by leveraging the disparity between estimated and (noisy) observed prices. Notably, this scheme does not necessitate a separate learning problem and instead we learn the parametrization while playing the game. We may show that every firm learns the true Nash-Cournot equilibrium strategy and the correct value of the misspecified parameter in an almost-sure sense.

Finally, we consider misspecified Markov decision problems in which transition matrices are unknown and the cost functions are misspecified. We propose three types of schemes: (1) misspecified value iteration; (2) misspecified policy iteration; and (3) misspecified Q-learning. The almost sure convergence and a non-asymptotic bound on the mean-squared error can be derived for the misspecified value iteration scheme. When the steplength is held constant, we may also get an optimized error bound for the averaged misspecified value funtion. Next, we propose a misspecified policy iteration scheme and provide an analogous asymptotic almost-sure convergence statement and an analysis of the rate of convergence. Finally, a constant steplength misspecified Q-learning scheme is presented and a suitable error bound based on iteration steps and steplength is provided.

References

[1] T. Hastie, R. Tibshirani, and J. H. Friedman, **The elements of statistical learning: data mining, inference, and prediction: with 200 full-color illustrations**. New York: Springer-Verlag, 2001.

[2] D. P. Bertsekas, A. Nedić, and A. E. Ozdaglar, **Convex analysis and optimization**. Athena Scientific, Belmont, MA, 2003.

[3] W. B. Powell, **Approximate Dynamic Programming: Solving the Curses of Dimensionality (Wiley Series in Probability and Statistics)**. Wiley-Interscience, 2007.

[4] D. P. Bertsekas, **Dynamic programming and optimal control. Vol. I**, 3rd ed. Athena Scientific, Belmont, MA, 2005.

[5] A. Ben-Tal, L. El Ghaoui, and A. Nemirovski, **Robust Optimization**, ser. Princeton Series in Applied Mathematics. Princeton University Press, October 2009.

[6] D. Bertsimas, D. B. Brown, and C. Caramanis, "Theory and applications of robust optimization," **SIAM Review**, vol. 53, no. 3, pp. 464–501, 2011.

[7] D. Bertsimas, V. Gupta, and N. Kallus, "Data-driven robust optimization," **arXiv : 1401.0212**.

[8] D. Goldfarb and G. Iyengar, "Robust portfolio selection problems," **Mathematics of Operations Research**, vol. 28, no. 1, pp. 1–38, 2003.

[9] D. Bertsimas and M. Sim, "The price of robustness," **Operations Research**, vol. 52, no. 1, pp. 35–53, 2004.

[10] D. Bertsimas and D. Pachamanova, "Robust multiperiod portfolio management in the presence of transaction costs," **Computers & Operations Research**, vol. 35, no. 1, pp. 3–17, 2008.

[11] R. Jiang, M. Zhang, G. Li, and Y. Guan, "Two-stage network constrained robust unit commitment problem," **European Journal of Operational Research**, vol. 234, no. 3, pp. 751–762, 2014.

[12] C. Zhao and Y. Guan, "Unified stochastic and robust unit commitment," **Power Systems, IEEE Transactions on**, vol. 28, no. 3, pp. 3353–3361, Aug 2013.

[13] R. Jiang, J. Wang, and Y. Guan, "Robust unit commitment with wind power and pumped storage hydro," **Power Systems, IEEE Transactions on**, vol. 27, no. 2, pp. 800–810, May 2012.

[14] C. Li and S. Liu, "A robust optimization approach to reduce the bullwhip effect of supply chains with vendor order placement lead time delays in an uncertain environment," **Applied Mathematical Modelling**, vol. 37, no. 3, pp. 707–718, 2013.

[15] E. Adida and G. Perakis, "Dynamic pricing and inventory control: robust vs. stochastic uncertainty models—a computational study," **Annals of Operations Research**, vol. 181, pp. 125–157, 2010.

[16] C.-T. See and M. Sim, "Robust approximation to multiperiod inventory management," **Operations Research**, vol. 58, no. 3, pp. 583–594, 2010, supplementary data available online.

[17] E. Adida and G. Perakis, "Dynamic pricing and inventory control: uncertainty and competition," **Operations Research**, vol. 58, no. 2, pp. 289–302, 2010.

[18] A. M. C. A. Koster, M. Kutschka, and C. Raack, "Robust network design: formulations, valid inequalities, and computations," **Networks**, vol. 61, no. 2, pp. 128–149, 2013.

[19] H. Hijazi, P. Bonami, and A. Ouorou, "Robust delay-constrained routing in telecommunications," **Annals of Operations Research**, vol. 206, pp. 163–181, 2013.

[20] F. Facchinei and J. S. Pang, **Finite-dimensional variational inequalities and complementarity problems. Vol. I**, ser. Springer Series in Operations Research. New York: Springer-Verlag, 2003.

[21] H. Jiang and H. Xu, "Stochastic approximation approaches to the stochastic variational inequality problem," **IEEE Transactions on Automatic Control**, vol. 53, pp. 1462–1475, 2008.

[22] J. Koshal, A. Nedic, and U. V. Shanbhag, "Regularized iterative stochastic approximation methods for stochastic variational inequality problems," **IEEE Transactions on Automatic Control**, vol. 58, no. 3, pp. 594–609, 2013.

[23] F. Yousefian, A. Nedić, and U. Shanbhag, "A regularized smoothing stochastic approximation (RSSA) algorithm for stochastic variational inequality problems," in **Proceedings of the Winter Simulation Conference (WSC)**, Dec 2013, pp. 933–944.

[24] A. Juditsky, A. Nemirovski, and C. Tauvel, "Solving variational inequalities with stochastic mirror-prox algorithm," **Stochastic Systems**, vol. 1, no. 1, pp. 17–58, 2011.

[25] F. Yousefian, A. Nedić, and U. V. Shanbhag, "Optimal robust smoothing extragradient algorithms for stochastic variational inequality problems," in **Proceedings of the IEEE Conference on Decision and Control (CDC)**, 2014.

[26] Y. Chen, G. Lan, and Y. Ouyang, "Accelerated schemes for a class of variational inequalities," **arXiv:1403.4164**.

[27] F. Facchinei and J.-S. Pang, **Finite-dimensional variational inequalities and complementarity problems. Vol. I**, ser. Springer Series in Operations Research. New York: Springer-Verlag, 2003.

[28] K. J. Astrom and B. Wittenmark, **Adaptive Control**, 2nd ed. Boston, MA, USA: Addison-Wesley Longman Publishing Co., Inc., 1994.

[29] R. Aguech, E. Moulines, and P. Priouret, "On a perturbation approach for the analysis of stochastic tracking algorithms." **SIAM Journal on Control and Optimization**, vol. 39, no. 3, pp. 872–899, 2000.

[30] L. Ljung and S. Gunnarsson, "Adaptation and tracking in system identification - a survey," **Automatica**, vol. 26, no. 1, pp. 7–21, 1990.

[31] H. Kushner, "Stochastic approximation: a survey," **Wiley Interdisciplinary Reviews: Computational Statistics**, vol. 2, no. 1, pp. 87–96, 2010.

[32] M. Uchiyama, "Formation of high-speed motion pattern of a mechanical arm by trial," **Transactions of the Society of Instrument and Control Engineers (Japan)**, vol. 14, pp. 706–712, 1978.

[33] S. Arimoto, S. Kawamura, and F. Miyazaki, "Formation of high-speed motion pattern of a mechanical arm by trial," **Journal of Robotic Systems**, vol. 1, pp. 123–140, 1984.

[34] K. L. Moore, **Iterative Learning Control for Deterministic Systems**, ser. Springer-Verlag Series on Advances in Industrial Control. London: Springer-Verlag, 1993.

[35] J. C. Gittins, **Multi-armed bandit allocation indices**. Wiley-Interscience Series in Systems and Optimization, Chichester: John Wiley & Sons, Ltd., 1989.

[36] M. N. Katehakis and J. A. F. Veinott, "The multi-armed bandit problem: Decomposition and computation," **Mathematics of Operations Research**, vol. 12, no. 2, pp. 262–268, 1987.

[37] P. Auer, N. Cesa-Bianchi, and P. Fischer, "Finite-time analysis of the multiarmed bandit problem," **Machine Learning**, vol. 47, no. 2-3, pp. 235–256, 2002.

[38] W. L. Cooper, T. Homem-de Mello, and A. J. Kleywegt, "Models of the spiral-down effect in revenue management," **Operations Research**, vol. 54, pp. 968–987, September 2006.

[39] B. T. Polyak, **Introduction to optimization**. New York: Optimization Software, Inc., 1987.

[40] F. Facchinei and J. S. Pang, **Finite-dimensional variational inequalities and complementarity problems. Vol. II**, ser. Springer Series in Operations Research. New York: Springer-Verlag, 2003.

[41] A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro, "Robust stochastic approximation approach to stochastic programming," **SIAM Journal on Optimization**, vol. 19, no. 4, pp. 1574–1609, Jan. 2009.

[42] A. Shapiro, D. Dentcheva, and A. Ruszczyński, **Lectures on stochastic programming**, ser. MP-S/SIAM Series on Optimization. Philadelphia, PA: SIAM, 2009, vol. 9, modeling and theory.

[43] B. T. Polyak and A. B. Juditsky, "Acceleration of stochastic approximation by averaging," **SIAM Journal on Control and Optimization**, vol. 30, no. 4, pp. 838–855, 1992.

[44] G. Lan, "An optimal method for stochastic composite optimization," **Mathematical Programming**, vol. 133, no. 1-2, pp. 365–397, Jun. 2012.

[45] S. Ghadimi and G. Lan, "Optimal stochastic approximation algorithms for strongly convex stochastic composite optimization I: A generic algorithmic framework," **SIAM Journal on Optimization**, vol. 22, no. 4, pp. 1469–1492, 2012.

[46] ——, "Optimal stochastic approximation algorithms for strongly convex stochastic composite optimization, II: shrinking procedures and optimal algorithms," **SIAM Journal on Optimization**, vol. 23, no. 4, pp. 2061–2089, 2013.

[47] M. Zinkevich, "Online convex programming and generalized infinitesimal gradient ascent," in **International Conference on Machine Learning (ICML)**, T. Fawcett and N. Mishra, Eds. AAAI Press, 2003, pp. 928–936.

[48] I. V. Konnov, **Equilibrium models and variational inequalities**, ser. Mathematics in Science and Engineering. Elsevier B. V., Amsterdam, 2007, vol. 210.

[49] A. Kannan and U. V. Shanbhag, "Distributed computation of equilibria in monotone Nash games via iterative regularization techniques," **SIAM Journal of Optimization**, vol. 22, no. 4, pp. 1177–1205, 2012.

[50] ——, "Distributed iterative regularization algorithms for monotone Nash games," **Proceedings of the IEEE Conference on Decision and Control (CDC)**, pp. 1963–1968, 2010.

[51] J. V. Burke and M. C. Ferris, "Weak sharp minima in mathematical programming," **SIAM Journal on Control and Optimization**, vol. 31, pp. 1340–1359, 1993.

[52] P. Marcotte and D. Zhu, "Weak sharp solutions of variational inequalities," **SIAM Journal on Optimization**, vol. 9, no. 1, pp. 179–189, 1999.

[53] M. C. Ferris and T. S. Munson, "Complementarity problems in GAMS and the PATH solver," **Journal of Economic Dynamics and Control**, vol. 24, no. 2, pp. 165–188, 2000.

[54] H. P. Young and S. Z. (eds), Eds., **Game theory and distributed control**, vol. 4. Elsevier, 20xx.

[55] N. Li and J. R. Marden, "Designing games to handle coupled constraints," in **Proceedings of the IEEE Conference on Decision and Control (CDC)**. IEEE, 2010, pp. 250–255.

[56] ——, "Designing games for distributed optimization," in **Proceedings of the IEEE Conference on Decision and Control (CDC)**, 2011, pp. 2434–2440.

[57] D. Fudenberg and D. K. Levine, **The theory of learning in games**, ser. MIT Press Series on Economic Learning and Social Evolution. Cambridge, MA: MIT Press, 1998, vol. 2.

[58] H. P. Young, **Strategic Learning and its Limits**. Oxford University Press, 2004.

[59] S. Hart, "Adaptive heuristics," **Econometrica**, vol. 73, no. 5, pp. 1401–1430, 2005.

[60] J. S. Shamma and G. Arslan, "Dynamic fictitious play, dynamic gradient play, and distributed convergence to Nash equilibria," **IEEE Transactions on Automatic Control**, vol. 50, no. 3, pp. 312–327, 2005.

[61] T. Başar, "Control and game-theoretic tools for communication networks," **Applied and Computational Mathematics**, vol. 6, no. 2, pp. 104–125, 2007.

[62] T. Alpcan and T. Başar, "Distributed algorithms for Nash equilibria of flow control games," **Annals of Dynamic Games**, vol. 7, 2003.

[63] Y. Pan and L. Pavel, "Games with coupled propagated constraints in optical network with multi-link topologies," **Automatica**, vol. 45, pp. 871–880, 2009.

[64] H. Yin, U. V. Shanbhag, and P. G. Mehta, "Nash equilibrium problems with scaled congestion costs and shared constraints," **IEEE Transactions on Automatic Control**, vol. 56, no. 7, pp. 1702–1708, 2011.

[65] F. Facchinei and J. S. Pang, "Nash Equilibria: The Variational Approach," **Convex Optimization in Signal Processing and Communication, Cambridge University Press**, 2009.

[66] G. Scutari and J.-S. Pang, "Joint sensing and power allocation in nonconvex cognitive radio games: Quasi-nash equilibria," in **Digital Signal Processing (DSP), 2011 17th International Conference on**. IEEE, 2011, pp. 1–8.

[67] A. P. Kirman, "Learning by firms about demand conditions," in **Adaptive economic models (Proc. Sympos., Math. Res. Center, Univ. Wisconsin, Madison, Wis., 1974)**. New York: Academic Press, 1975, pp. 137–156. Math. Res. Center, Univ. Wisconsin, Publ. No. 34.

[68] G. I. Bischi, A. Naimzada, and L. Sbragia, "Oligopoly games with local monopolistic approximation," **Journal of Economic Behavior and Organization**, vol. 62, pp. 371–388, 2007.

[69] G. I. Bischi, L. Sbragia, and F. Szidarovszky, "Learning the demand function in a repeated Cournot oligopoly game," **International Journal of Systems Science**, vol. 39, no. 4, pp. 403–419, 2008.

[70] F. Szidarovszky, "Global stability analysis of a special learning process in dynamic oligopolies," **Journal of Economic and Social Research**, vol. 9, pp. 175–190, 2004.

[71] F. Szidarovszky and J. B. Krawczyk, "On stable learning in dynamic oligopolies," **Pure Mathematics and Applications**, vol. 15, no. 4, pp. 453–468, 2004.

[72] D. Léonard and K. Nishimura, "Nonlinear dynamics in the Cournot model without full information," **Annals of Operations Research**, vol. 89, pp. 165–173, 1999, nonlinear dynamical systems and adaptive methods (Vienna, 1997).

[73] B. Hobbs, "Linear complementarity models of Nash-Cournot competition in bilateral and poolco power markets," **IEEE Transactions on Power Systems**, vol. 16, no. 2, pp. 194–202, 2001.

[74] B. F. Hobbs and J. S. Pang, "Nash-Cournot equilibria in electric power markets with piecewise linear demand functions and joint constraints," **Operations Research**, vol. 55, no. 1, pp. 113–127, 2007.

[75] L. Pavel, "A noncooperative game approach to OSNR optimization in optical networks," **IEEE Transactions on Automatic Control**, vol. 51, no. 5, pp. 848–852, 2006.

[76] C. H. Papadimitriou and M. Yannakakis, "On bounded rationality and computational complexity," Indiana University, Tech. Rep., 1994.

[77] H. A. Simon, **The sciences of the artificial (3rd ed.)**. Cambridge, MA, USA: MIT Press, 1996.

[78] G. Scutari and J.-S. Pang, "Joint sensing and power allocation in nonconvex cognitive radio games: Nash equilibria and distributed algorithms," **Information Theory, IEEE Transactions on**, vol. 59, no. 7, pp. 4626–4661, July 2013.

[79] H. Jiang and U. Shanbhag, "On the solution of stochastic optimization problems in imperfect information regimes," in **Simulation Conference (WSC), 2013 Winter**, Dec 2013, pp. 821–832.

[80] V. S. Borkar, **Stochastic Approximation: A Dynamical Systems Viewpoint**. Cambridge University Press, 2008.

[81] H. J. Kushner and G. G. Yin, **Stochastic approximation and recursive algorithms and applications**, 2nd ed., ser. Applications of Mathematics (New York). New York: Springer-Verlag, 2003, vol. 35, stochastic Modelling and Applied Probability.

[82] J. Hofbauer and W. H. Sandholm, "Stable games and their dynamics," **Journal of Economic Theory**, vol. 144, no. 4, pp. 1665–1693, 2009.

[83] M. Fox and J. Shamma, "Population games, stable games, and passivity," in **Decision and Control (CDC), 2012 IEEE 51st Annual Conference on**, Dec 2012, pp. 7445–7450.

[84] G. I. Bischi, C. Chiarella, M. Kopel, and F. Szidarovszky, **Nonlinear oligopolies**. Berlin: Springer-Verlag, 2010, stability and bifurcations.

[85] R. J. Aumann, "Agreeing to disagree," **The Annals of Statistics**, vol. 4, no. 6, pp. pp. 1236–1239, 1976.

[86] J. Littlewood, **Mathematical Miscellany**, B. Bollabos, Ed., 1953.

[87] T. Shelling, **The Strategy of Conflict**. Harvard University Press, Cambridge, Massachusetts, 1960.

[88] M. A. Crew and D. Parker, Eds., **International Handbook on Economic Regulation**. Edward Elgar, 2006.

[89] S. Dafermos, "Sensitivity analysis in variational inequalities," **Mathematics of Operations Research**, vol. 13, no. 3, pp. 421–434, 1988.

[90] T. W. Anderson and J. B. Taylor, "Strong consistency of least squares estimates in dynamic models," **The Annals of Statistics**, vol. 7, no. 3, pp. 484–489, 1979.

[91] R. Bellman, **Dynamic Programming**, 1st ed. Princeton, NJ, USA: Princeton University Press, 1957.

[92] T. W. Anderson and L. A. Goodman, "Statistical inference about markov chains," **Annals of Mathematical Statistics**, vol. 28, pp. 89–110, 1957.

[93] L. Ljung, "System identification: Theory for the user," **Prentice Hall Inf and System Sciencess Series, New Jersey**, vol. 7632, 1987.

[94] F. Han and H. Liu, "Transition matrix estimation in high dimensional time series," in **Proceedings of the 30th International Conference on Machine Learning (ICML-13)**, S. Dasgupta and D. Mcallester, Eds., vol. 28, no. 2. JMLR Workshop and Conference Proceedings, May 2013, pp. 172–180.

[95] A. Nilim and L. El Ghaoui, "Robust control of Markov decision processes with uncertain transition matrices," **Operations Research**, vol. 53, no. 5, pp. 780–798, September-October 2005.

[96] E. Delage and S. Mannor, "Percentile optimization for markov decision processes with parameter uncertainty," **Operations Research**, vol. 58, no. 1, pp. 203–213, 2010.

[97] C. J. C. H. Watkins and P. Dayan, "Technical note Q-learning," **Machine Learning**, vol. 8, pp. 279–292, 1992.

[98] J. N. Tsitsiklis and R. Sutton, "Asynchronous stochastic approximation and Q-learning," in **Machine Learning**, 1994, pp. 185–202.

[99] H. Chang, J. Hu, M. Fu, and S. Marcus, **Simulation-Based Algorithms for Markov Decision Processes**, ser. Communications and Control Engineering. Springer London, 2013.

[100] R. I. Brafman and M. Tennenholtz, "R-max - a general polynomial time algorithm for near-optimal reinforcement learning," **Journal of Machine Learning Research**, vol. 3, pp. 213–231, Mar. 2003.

[101] P. L. Bartlett and A. Tewari, "Regal: A regularization based algorithm for reinforcement learning in weakly communicating mdps," in **Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence**, ser. UAI '09. Arlington, Virginia, United States: AUAI Press, 2009, pp. 35–42.

[102] P. Auer, T. Jaksch, and R. Ortner, "Near-optimal regret bounds for reinforcement learning," in **Advances in Neural Information Processing Systems 21**, D. Koller, D. Schuurmans, Y. Bengio, and L. Bottou, Eds. Curran Associates, Inc., 2009, pp. 89–96.

[103] H. Jiang and U. V. Shanbhag, "On the solution of stochastic optimization problems in imperfect information regimes," in **Winter Simulation Conference**, 2013, pp. 821–832.

[104] R. Howard, **Dynamic Programming and Markov Processes**. The MIT press, New York London, 1960.

[105] D. P. Bertsekas, **Dynamic Programming and Optimal Control, Vol. II**, 3rd ed. Athena Scientific, 2007.

[106] A. Gosavi, "Boundedness of iterates in Q-learning," **Systems & Control Letters**, vol. 55, no. 4, pp. 347–349, 2006.

[107] C. L. Beck and R. Srikant, "Error bounds for constant step-size Q-learning," **Systems & Control Letters**, vol. 61, no. 12, pp. 1203 – 1208, 2012.